STRUCTURAL BIOLOGY OF *LEGIONELLA PNEUMOPHILA* EFFECTORS

A Thesis Submitted to the College of

Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

KEVIN ANDREW VOTH

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use, which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make use of the material in this thesis in whole or in part should be addressed to:

Head of the Department of Biochemistry
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan, S7N 5C9
Canada

ABSTRACT

Legionella pneumophila is a Gram-negative intracellular pathogen that causes Legionnaires' disease and Pontiac fever in elderly or immunocompromised humans. The ability of *Legionella* to thrive within host cells depends on the *Legionella*-containing vacuole (LCV) which, in turn, relies on the activity of secreted effector proteins for its formation. Effectors are highly variable in structure and function, and functional redundancy is prevalent among them. Consequently, relating structural data to function provides an attractive avenue of research into molecules which are unlikely to exhibit a phenotype upon gene deletion. Our lab relies on X-ray crystallography for macromolecular structure determination. Structural data may point to a function for the protein of interest, which can be verified using mutagenesis, biochemical assays or some combination thereof. This dissertation explores the structure and putative function of effectors LpnE(*lpg2222*), MavE(*lpg2344*) and MavL(*lpg2526*).

LpnE (*Legionella pneumophila* Entry) is a Sel1-like repeat (SLR) protein implicated in host cell invasion. During infection, a eukaryotic polyphosphate 5-phosphatase, called Oculocerebrorenal syndrome of Lowe protein 1 (OCRL1), is recruited to the LCV by an interaction with LpnE and restricts bacterial replication by an unknown mechanism. The crystal structure of His-LpnE(73-375) reveals a typical SLR super-helix with a concave surface implicated in protein-protein interactions. Herein, critical residues promoting the LpnE-OCRL interaction are uncovered using size exclusion chromatography with multi-angle light scattering (SEC-MALS). In addition, we show that LpnE localizes to *cis*-Golgi using its signal peptide. These findings are compiled into a mechanistic hypothesis where: (1) LpnE localizes to the LCV by its predicted signal peptide. (2) OCRL binding prevents liberation of LpnE from the LCV and (3) renders LpnE unable to promote infection by mediating protein-protein interactions in the cytosol.

MavE is one of many proteins identified as a secreted effector based on its ability to rescue LCV localization of a translocation deficient SidC (SidCΔ100). Our collaborator, Dr. Yousef Abu-Kwaik, has obtained a unique phenotype for Δ*lpg2344* (MavE) mutants, in which the LCV fuses with lysosomes (unpublished data). He suggests that MavE interacts with proteins harbouring phosphotyrosine-binding domains (PTBs) using its NPxY motif. The recruitment of these binding

partners may impede autophagic trafficking. The crystal structure of MavE(39-172) presented in this dissertation has an overall structure reminiscent of the grass pollen allergen, Phlp 5b, with the NPxY motif located on a loop of poorly defined electron density. This loop has no counterpart in Phlp 5b and has flexibility that may accommodate protein-protein interactions. These structural data corroborate the proposed role of the NPxY motif while revealing a scaffold domain previously seen in the grass pollen allergen, Phlp 5b.

MavL is another secreted effector identified in the same manner as MavE. Presently, there is little published data available on the function of MavL. Elizabeth Hartland et al. have found by yeast two-hybrid that an E2 ubiquitin conjugating enzyme called Ube2q1 interacts directly with MavL, although we were unable to reproduce this interaction *in vitro*. The crystal structure of MavL(42-435) reveals an ADP-ribose binding macrodomain with homology to those that recognize mono-ADP-ribosylated targets. We confirmed the interaction of MavL and ADP-ribose by isothermal titration calorimetry (ITC), giving a dissociation constant of 13μM. Intriguingly, MavL contains a pair of neighboring aspartate residues in the same location as the catalytic glutamates of poly-ADP-ribose glycohydrolase (PARG) enzymes. We propose that MavL exhibits either ADP-ribose reader or eraser activity. Further studies are needed to investigate the role of ADP-ribosylation in MavL functionality.

ACKNOWLEDGEMENTS

I would first like to thank my supervisor, Dr. Miroslaw (Mirek) Cygler, for his guidance, patience and support throughout this work. Mirek is an excellent teacher and always has the perfect analogy to explain an abstract concept. He was available for late night calls when I was struggling with data processing and met with me weekly to discuss my progress in the lab. Knowing that my work was a priority for him gave me a sense of confidence. Mirek also encouraged me to think for myself and develop independence. I could not have asked for a better supervisor.

I am grateful for my committee members, Dr. Yu Luo, Dr. Ron Geyer, Dr. Adelaine Leung and Dr. Jeremy Lee for their help in keeping me focused and on-track. All members of the Cygler Lab that I have known during my studies deserve recognition for their scientific advice and friendship: Dr. Michal Boniecki, Dr. ThirumalaiSelvi Ulaganathan, Dr. Karin van Straaten, Dr. Maia Cherney, Dr. Sergei Kalynych, Dr. Andrey Grishin, Dr. Usha Nair, Dr. Lei Li, Dr. Ivy Yeuk Wah Chung, Dr. Ksenia Beyrakhova, Anita Abglor, Ben Flath and Caishuang Xu. The Geyer Lab also deserves recognition for helping me learn the technique of phage display. I would like to extend my thanks to the Protein Characterization and Crystallization Facility (PCCF) and the Canadian Light Source – Canadian Macromolecular Crystallography Facility (CLS-CMCF) beamline scientists for their expertise and equipment. The Biochemistry Department at the University of Saskatchewan has also been very helpful to me.

I would like to thank my family for their support and encouragement throughout this process. My brothers, Ryan and Jamie, who I can happily call my friends, have been a source of humor, lightheartedness and wisdom. My partner, Bonnie Chu, has been a joy to share life with during my final years of study. Finally, I dedicate this dissertation to my parents, who have always encouraged me to pursue my dreams.

Table of Contents

List of Tables

Chapter 5

Chapter 6

Chapter 7

List of Figures

Chapter 2

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

List of Abbreviations

ADP	Adenosine diphosphate
ADPR	ADP-ribosylated
AHL	α-helical lobe
ATP	Adenosine triphosphate
BSA	Bovine serum albumin
cAMP	Cyclic adenosine monophosphate
CAPTPase	Cold-active protein tyrosine phosphatase
cGMP	Cyclic guanosine monophosphate
Cryo-EM	Single particle electron cryomicroscopy
DLS	Dynamic light scattering
DNA	Deoxyribonucleic acid
Dot/Icm	Defect in organelle transport / Intracellular multiplication
DUB	Deubiquitinase
EEA1	Early endosome antigen 1
ER	Endoplasmic reticulum
ERC	Endocytic recycling compartment
ExPEC	Extra-intestinal pathogenic *E. coli*
FPLC	Fast protein liquid chromatography
GAP	GTPase activating protein
GDI	GDP-dissociation inhibitor
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GFP	Green fluorescent protein

GTP	Guanosine triphosphate
HAD	Haloacid dehalogenase
HCl	Hydrochloric acid
Hcp	*H. pylori* cysteine-rich protein
HGPRT	Hypoxanthine-guanine phosphoribosyltransferase
ITC	Isothermal titration calorimetry
LAMP1	Lysosome-associated membrane protein 1
LBS	Lipid binding site
LC3	microtubule-associated protein 1A/1B light chain 3B
LCV	*Legionella*-containing vacuole
LIC	Ligation-independent cloning
LpnE	*L. pneumophila* Entry protein
LRR	Leucine-rich repeat
MAD	Multiple anomalous dispersion
mART	mono-ADP-ribosyltransferase
Mav	More proteins associated with vacuole colocalization
MDP1	Magnesium-dependent phosphatase
NaCl	Sodium chloride
NAD	Nicotinamide adenine dinucleotide
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
OCRL	Oculocerebrorenal syndrome of Lowe
OMCC	Outer membrane core complex
PARG	Poly-ADP-ribose glycohydrolase

PDE	Phosphodiesterase
PE	Phosphatidylethanomlamine
PEG	Polyethylene glycol
PH	Pleckstrin homology
PI(x,y,z)P3	Phosphatidylinositol(x,y,z)triphosphate
PIxK	Phosphatidylinositol-x-kinase
PP	Pancreatic polypeptide
PPM	Metal-dependent protein phosphatases
PPP	Phosphoprotein phosphatase
PsiBLAST	Position-specific iterated Basic Local Alignment Search Tool
PTB	Phosphotyrosine binding
Rav	Regions allowing vacuole colocalization
RMSD	Root mean square deviation
RNA	Ribonucleic acid
SAD	Single anomalous dispersion
SCF	Skp1-cullin-F box
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC-MALS	Size exclusion chromatography with multi-angle light scattering
SLR	Sel1-like repeat
SNARE	Soluble N-ethylmaleimide-sensitive factor
T4SS	Type 4 Secretion System
TEV	Tobacco etch virus protease
TRP	Tetratricopeptide repeat

1. Introduction

Legionella pneumophila is the causative agent of Legionnaire's disease and a milder illness known as Pontiac fever. It is a Gram-negative intracellular pathogen that infects a large spectrum of freshwater amoeba as its primary hosts. Once transmitted to the human lung by contaminated aerosols, *L. pneumophila* replicates within alveolar macrophages and gives rise to disease in elderly or immunocompromised individuals. The ability of *L. pneumophila* to replicate in human cells stems from its strategy for infecting lower eukaryotes, but *Legionella* does not efficiently mediate disease transmission in humans. Consequently, *L. pneumophila* has been named an "accidental pathogen" and humans are a dead-end host for this organism.

L. pneumophila depends on successful formation and maturation of the so-called "*Legionella*-containing vacuole" (LCV) for survival within host cells. This compartment is derived initially from the host cell membrane, but later acquires components of the endoplasmic reticulum (ER) and mitochondrial membranes. Formation of the LCV depends on an arsenal of effector proteins, which are secreted by a membrane-spanning complex called the Type IVB (Dot/Icm) secretion system. These secreted effectors are highly variable in structure and function, and hijack host cell processes at many levels: they can act as enzymes or transcription factors, modulate small GTPase activity and interfere with ubiquitination, cytoskeletal organization and even the function of other effectors. Each *L. pneumophila* strain encodes a unique complement of over 300 effectors. Consequently, the total number of *Legionella* effectors to be studied far exceeds this number. The most investigated strains are Philadelphia and Paris, each secreting ~330 effectors. Redundancy in the function of effector proteins often precludes their analysis by knockout studies, rendering structural biology an attractive avenue for investigating the role of effectors as infective agents.

Our lab relies on X-ray crystallography to obtain structural data on effectors. These data may point to a function for the protein of interest, which can be supported using biochemical assays. In this way, a functional understanding of effectors can be derived, in part, from structural knowledge. To this end, I have structurally characterized three effectors encoded by *L. pneumophila* (sp. Philadelphia): LpnE, MavE and MavL.

2. Review of the Literature

2.1. *Legionella pneumophila* background

2.1.1. Uptake into Host Cells

Transmissive *L. pneumophila* is taken into the host cell by macropinocytosis, a process in which membrane projections fuse to generate a vesicle called the macropinosome. These projections or "ruffles" arise from the rearrangement of cortical actin filaments. It has been found that phosphatidylinositol(3,4,5)triphosphate (PI(3,4,5)P3) and PI(3,4)P2 play critical roles in micropinocytosis (de Carvalho et al., 2015). As such, enzymes involved in the generation of these phospholipids, such as phosphatidylinositol-3-kinase (PI3K) contribute to macropinosome formation. Indeed, PI3K converts PI(4,5)P2 to PI(3,4,5)P3 at macropinocytic entry sites. The role of PI(3,4,5)P3 and PI(3,4)P2 in micropinocytosis has been linked to actin polymerization. For example, PI(3,4,5)P3 recruits GEFs and GAPs harbouring PI(3,4,5)P3 PH domains, which direct actin polymerization by RhoGTPase activity. After internalization of the macropinosome, PI(4,5)P2 is regenerated on the cell surface (Steiner et al., 2017).

The internalization process of *L. pneumophila* is thought to be initiated by the deposition of a secreted effector into the host cell cytosol (Watarai et al., 2001). Although the exact mechanism of host cell invasion remains to be elucidated, structural genes *rtxA* and *enhC* have been found to play a role in attachment. The *rtxA* gene product, RtxA, contains eight RTX motifs, which facilitate Ca^{2+} binding and host cell attachment. Meanwhile, *enhC* encodes a Sel1-like repeat (SLR) protein that mediates protein-protein interactions (Cirillo et al., 2000). LpnE is another SLR protein encoded by *L. pneumophila* that compensates for defects in entry arising from *enhC* deletion (Bandyopadhyay et al., 2012). Although the role of SLR proteins in host cell entry remains unclear, protein-protein interactions typically occur at the concave surface of the SLR super-helix (Cliff et al., 2006; Das et al., 1998; Scheufler et al., 2000). Thus, determining eukaryotic binding partners of LpnE/EnhC may provide mechanistic insights into how these proteins facilitate entry.

2.1.2. Biphasic Life Cycle of *Legionella pneumophila*

The life cycle of *L. pneumophila* can be called biphasic, as this bacterium exists in either a replicative or transmissive form. During its growth within host cells the replicative form is adopted, and bacteria are non-motile, slender rods, which lack cytotoxicity. When nutrient levels become limiting, *L. pneumophila* differentiate into a morphologically distinct, flagellated and virulent transmissive form. It is this transmissive form that escapes from the spent host cell and reinitiates the process of infection. These distinct lifeforms allow *L. pneumophila* to conserve energy by expressing the genes for virulence and transmission only when conditions for growth and replication are no longer being met (Oliva et al., 2018).

2.1.3. The *Legionella pneumophila* Type IV Secretion System

The Type IV Secretion System (T4SS) is a functionally diverse nanomachine that shares evolutionary roots with the bacterial conjugation apparatus. These envelope-spanning multiprotein complexes translocate a wide variety of substrates out of the cell, including DNA, toxins and biofilm components. In the case of *L. pneumophila*, the T4SS is a dedicated effector-translocating machine (Grohmann et al., 2017). There are two phylogenetic families of T4SS in Gram-negative bacteria, designated IVA and IVB. The IVA system is often represented by the *Agrobacterium tumefaciens* VirB/VirD4 T4SS, which is comprised of 11 VirB proteins and one VirD4 subunit (Li and Christie, 2018). These proteins have been functionally classified as cytoplasmic ATPases, inner membrane platform components, outer membrane core complex and pilus assembly components. The Dot/Icm secretion system encoded by *L. pneumophila* is the IVB paradigm and contains 27 subunits, many of which lack homologues in the IVA system (Grohmann et al., 2017). These subunits combine to form the wheel, collar, cylinder and cytoplasmic complexes that make up the T4BSS (Figure 2.1).

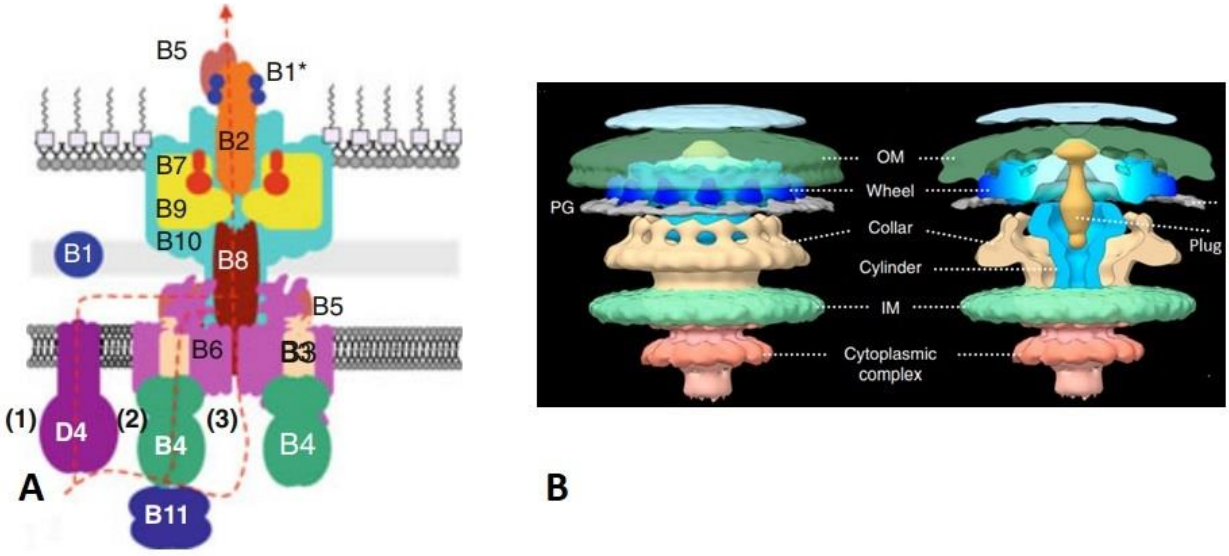

Figure 2. 1 Schematic of type IV secretion systems (T4SS). (A) Diagram of A. tumefaciens T4ASS showing 11 VirB subunits and the VirD4 chaperone. (B) Three-dimensional surface rendering of the L. pneumophila T4BSS revealed by cryo-electron tomography and subtomogram averaging. Components of the system are designated in white. PG = peptidoglycan, OM = outer membrane, IM = inner membrane. Figure adapted from (Chetrit et al., 2018; Li and Christie, 2018).

A structure of the *L. pneumophila* IVB system has been obtained *in situ* by electron cryotomography. This model shows striking similarity to the IVA systems reported previously, suggesting a shared mechanism of translocation despite little similarity in the sequence and subunits of these complexes (Ghosal et al., 2017). Indeed, only the hat domain of the outer membrane core complex (OMCC) is structurally equivalent in both IVA and IVB systems. The proteins comprising these hat domains are DotG and VirD10 in *L. pneumophila* and *A. tumefaciens*, respectively.

2.1.4. Phosphoinositide conversion on the *Legionella*-containing vacuole (LCV)

PI(3,4,5)P3 and PI(3,4)P2 direct actin polymerization at the site of micropinocytosis. Consequently, the engulfed macropinosome is initially rich in PI(3,4,5)P3 and PI(3,4)P2. Upon internalization, these phospholipids are traded for PI3P in an Icm/Dot-independent manner.

Rab5 regulates the appearance and disappearance of PI3P in eukaryotic cells by recruiting PI3-kinases or PI4- and PI5-phosphatases, respectively (Shin et al., 2005). Since the disappearance of PI(3,4,5)P3 and PI(3,4)P2 is correlated with the formation of PI3P, Rab5-recruited phosphatases may be involved in this process. In vesicles harboring *Legionella* (LCV) containing an intact Icm/Dot system, PI3P is exchanged for PI4P in an Icm/Dot-dependent manner (Steiner et al., 2017).

The conversion of PI3P to PI4P in non-infected cells depends on host kinases. For example, the phosphatidylinositol is phosphorylated at position 4 by phosphatidylinositol-4-kinase (PI4K) enzymes, which are divided into subtypes based on the subcellular compartment they are associated with. In cells infected with *Legionella*, the small GTPase Arf1 localizes to the LCV and recruits PI4K IIIβ, an enzyme that typically binds to Golgi and the nucleus. It is therefore possible that PI4K IIIβ plays a role in generating PI4P on the LCV, although direct evidence for the association of PI4K IIIβ with the LCV has yet to be obtained (Steiner et al., 2017).

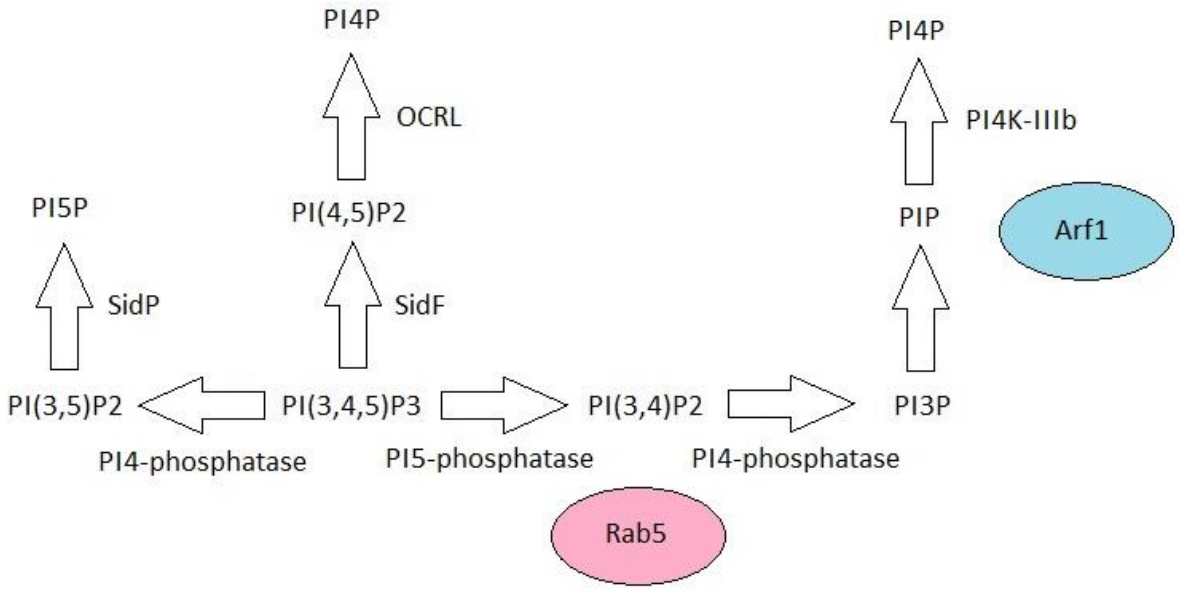

Figure 2. 2 Schematic of phosphoinositide conversion on the Legionella-containing vacuole (LCV). Rab5 (pink) and Arf1 (blue) recruit PI4/5-phosphatases and PI4K IIIβ to the LCV, where they generate PI3P and, ultimately, PI4P from PI(3,4,5) initially present on the macropinosome.

The Icm/Dot-dependent disappearance of PI3P from the LCV is also caused by effectors with PI metabolizing activity, such as SidF, LepB and SidP. LepB is a PI-4-kinase that acts on PI3P to generate PI(3,4)P2 which, in turn, is converted to PI4P by the 3-phosphatase SidF (Dong et al., 2016). Another substrate of SidF is PI(3,4,5)P3, a dominant lipid on the early macropinosome. This lipid is changed to PI(4,5)P2 by the 3-phosphatase activity of SidF (Hsu et al., 2012). The eukaryotic inositol polyphosphate 5-phosphatase, OCRL, then acts on PI(4,5)P2 to generate PI4P, thereby completing the conversion of PI(3,4,5)P3 to PI4P on the LCV (Zhang et al., 1995). SidP also promotes the loss of PI3P from the LCV, although it does not lead directly to the formation of PI4P. Instead, SidP uses its 3-phosphatase activity to generate PIP and PI5P from PI3P and PI(3,5)P2, respectively (Toulabi et al., 2013). Thus, *Legionella* has systems in place to convert PI3P, PI(3,4,5)P3 and PI(3,5)P2 to PI4P and PI5P on the LCV, while promoting PI4P production through the phosphorylation of PIP by PI4K IIIβ (Figure 2.2).

2.1.5. Effectors that alter the host phosphoproteome

L. pneumophila causes significant changes in the host cell phosphoproteome by secreting effectors that act as kinases and phosphatases. Several *L. pneumophila* kinases share sequence similarity with eukaryotic kinases, suggesting potential acquisition by horizontal gene transfer (Garcia-Vallve et al., 2003). Effectors acting as phosphatases have also been found to mimic eukaryotes by adopting haloacid dehalogenase-like (HAD-like) (Quaile et al., 2018) or phosphotyrosine phosphatase domains (Beyrakhova et al., 2018; Jia et al., 2018). Since phosphorylation plays a critical role in various host signaling processes, an understanding of how this post-translational modification is controlled by *L. pneumophila* is essential. The following is an overview of recent research into *Legionella* kinases and phosphatases.

The ‘eukaryotic-like’ kinases of *L. pneumophila* include LegK1, LegK2, LegK3, LegK4 and LegK7. These effectors phosphorylate serine or threonine residues on target proteins. Specifically, LegK1 phosphorylates IκB and upregulates the NF-κB pathway. The targets of LegK3 and LegK4 are unknown (Grishin et al., 2015), although recent structural data have provided insight into LegK4 functionality (see section 2.2.2.5).

In addition to acquiring kinases from eukaryotic genomes, bacteria have evolved their own set of kinases with structural similarity to nonregulatory kinases (Fookes et al., 2011; Hemrajani et al., 2010; Kim et al., 2005). Features that distinguish bacterial kinases from nonregulatory kinases are their small kinase domain and the absence of an activation loop. Indeed, the classification of LegK2 as 'eukaryotic-like' has been argued on the basis that an activation loop is absent in this molecule (Hervet et al., 2011).

LegK2 promotes the virulence of *L. pneumophila* by phosphorylating ARPC1B and ARP3 subunits of the ARP2/3 complex, leading to a local interference of actin remodeling around the LCV. Thus, the activity of LegK2 allows the LCV to evade endocytic trafficking (Michard et al., 2015). Another *Legionella* kinase, LegK7, was found to mimic the host Hippo kinase and phosphorylate MOB1. In so doing, transcriptional regulators TAZ and ZAP1 are degraded, leading to altered expression of genes associated with the PPARγ transcription factor (Lee and Machner, 2018).

Several phosphatases have been identified in the *L. pneumophila* genome. The effector Lem4 is a phosphotyrosine phosphatase with possible activity toward CSF1R based on its preference for the consensus motif of this receptor (Beyrakhova et al., 2018). CSF1R resides on the LCV during infection (Hoffmann et al., 2014) and activates the mitogen-activated protein kinase / extracellular signal-regulated kinase kinase and phosphatidylinositol-3-kinase pathways. These signaling pathways depend on dimerization and autophosphorylation of CSF1R at Tyr708 and Tyr809. Dephosphorylation of CSF1R by Lem4 would therefore prevent downstream signaling events, although the presence of such activity during infection has yet to be reported. A crystal structure of the N-terminal Lem4 haloacid dehalogenase-like (HAD-like) domain showed structural homology with mouse magnesium-dependent phosphatase (MDP1). Consistent with this finding, Lem4 has specificity toward phenyl-phosphate containing substrates, such as pNPP and pTyr (Beyrakhova et al., 2018).

WipA and WipB are two other bacterial phosphatases secreted by *L. pneumophila*. Both proteins harbor a serine/threonine phosphatase domain resembling that seen in the eukaryotic phosphoprotein phosphatase (PPP) family (Jia et al., 2018; Prevost et al., 2017). WipB localizes to LAMP1-positive lysosomal compartments and binds the v-ATPase and associated LAMTOR1 phosphoprotein (Prevost et al., 2017). Another *L. pneumophila* effector, SidK, has

also been found to inhibit the v-ATPase and prevent acidification of the LCV (Zhao et al., 2017). Since neither *wipB* nor *sidK* mutants caused a defect in the intracellular replication of *L. pneumophila*, these effectors may act redundantly to prevent LCV acidification by the v-ATPase (Prevost et al., 2017). WipA is an unconventional protein tyrosine phosphatase (PTP) with structural resemblance to cold-active protein tyrosine phosphatase (CAPTPase). The substrates of WipA have yet to be determined (Jia et al., 2018).

2.1.6. Modulation of the small GTPase Rab1 by *Legionella* effectors

Rabs belong to the Ras-like small GTPase superfamily. These proteins are involved in vesicle fusion, budding, tethering and transport. To influence the trafficking of intracellular compartments, Rabs engage in a process of membrane binding and dissociation. Briefly, Rabs are prenylated by an enzyme called geranylgeranyltransferase, which promotes membrane localization. This is followed by Rab activation with a guanine nucleotide exchange factor (GEF), that displaces bound GDP with GTP. Nucleotide exchange causes two small areas of the Rab protein, called switch I and switch II regions, to transition from disordered to ordered states (Luitz et al., 2016). Structural organization of these switch regions propels activated Rab to adopt a conformation that interacts with specific effector proteins, such as tethering factors or motor proteins involved in transport. Eventually, the biological role of the Rab is complete and a GTPase activating protein (GAP) hydrolyzes bound GTP to GDP, deactivating the Rab. At this point, a GDP-dissociation inhibitor (GDI) binds GDP-Rab and extracts it from the membrane (Figure 2.3). GDI-GDP-Rab exists as a soluble complex until another cycle of membrane targeting begins (Goody et al., 2017).

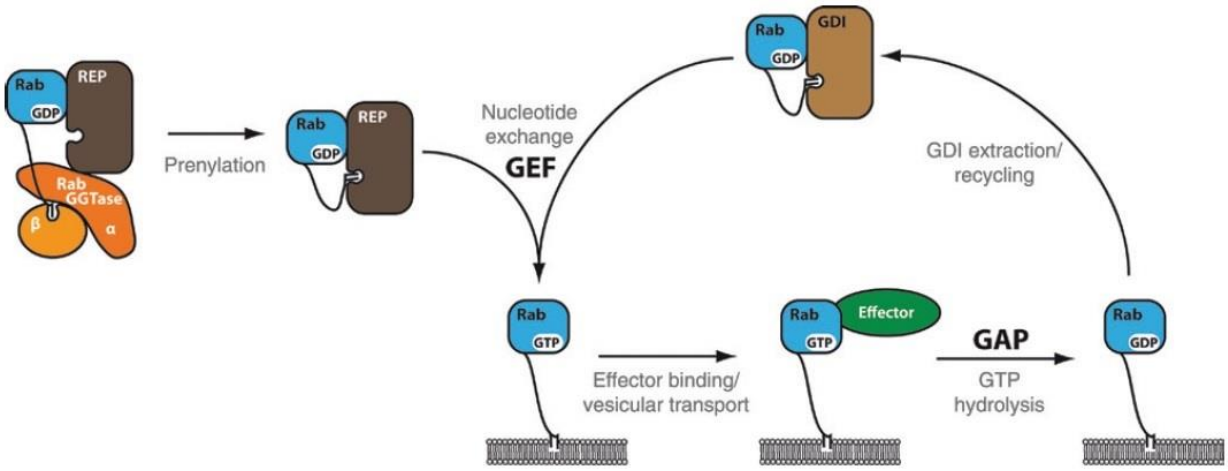

Figure 2. 3 Regulation of Rab protein activity by Rab-escort protein (REP), prenylation, guanine nucleotide exchange factor (GEF), GTPase activating protein (GAP) and GDP-dissociation inhibitor. The GTP-bound Rab resists extraction until it is deactivated by GTP hydrolysis. Figure adapted from (Goody et al., 2017).

Importantly, GEFs target Rabs to membranes. Since GDIs have lower affinity for activated than deactivated Rabs, the activation of a membrane bound Rab by GEF allows it to resist extraction from that membrane by GDI. Conversely, GAPs effectively remove Rabs from their resident membranes (Goody et al., 2017). These properties are exploited by *Legionella* effectors in the modulation of Rab activity. For example, by producing effectors that covalently modify Rabs such that the active conformation persists despite GTP hydrolysis, a Rab can remain membrane-associated for long periods of time. Moreover, host GAPs are often unable to deactivate this modified Rab (Spanò and Galán, 2018). What follows is a discussion of the methods used by *L. pneumophila* to modulate the activity of Rab1, a main regulator of the secretory pathway that is targeted by six different *L. pneumophila* effectors (Qiu and Luo, 2017a; Steiner et al., 2017).

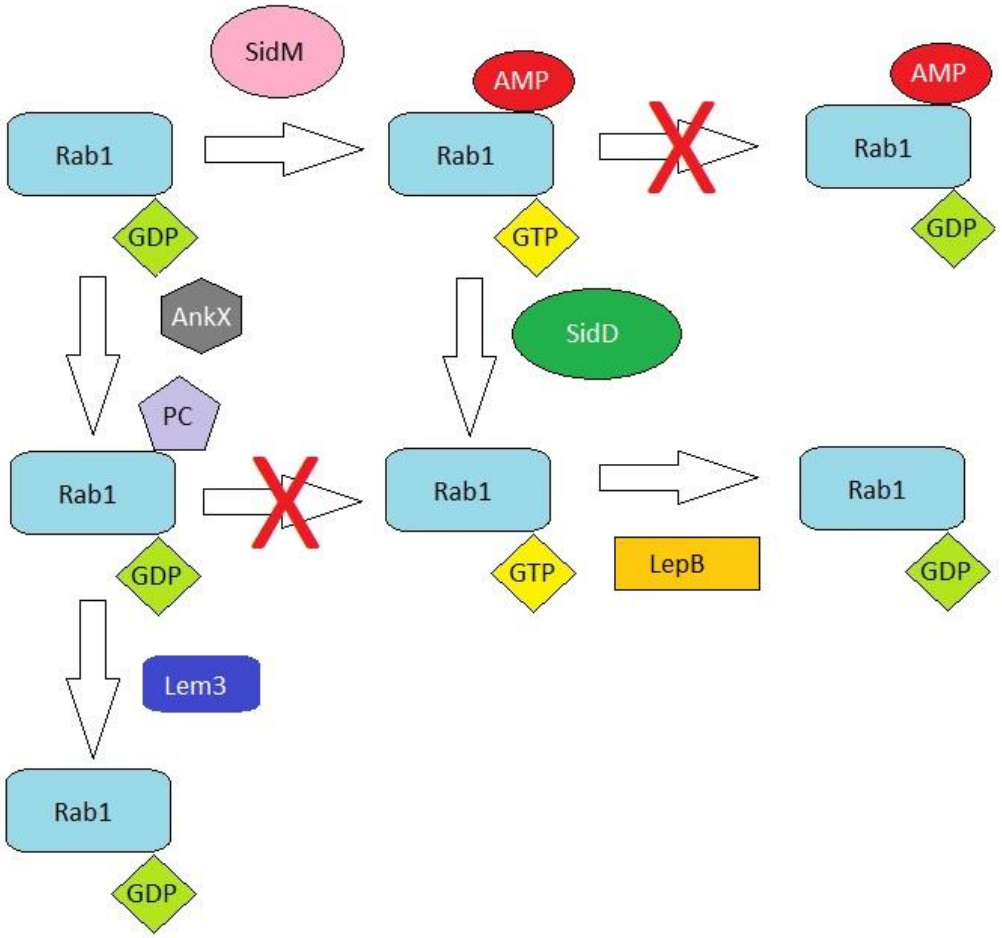

Figure 2. 4 Schematic of Rab1 regulation by Legionella pneumophila effectors. SidM (pink) promotes activation of Rab1 (light blue) while preventing canonical deactivation by host GAPs. SidD (dark green) initiates the process of deactivation by deAMPylating Rab1, rendering it a substrate of LepB (orange) or host GAPs. Alternatively, Rab1 can be retained in an inactive state by AnkX (grey), which PCylates the protein. This process is reversed by Lem3 (blue).

When localized to the LCV, Rab1 drives the recruitment of ER-derived vesicles to the pathogen vacuole. This process is a vital part of LCV maturation and in the establishment of a replicative niche. It is important to note that not all six effectors targeting Rab1 promote its activation. In fact, some exist to reverse the activating effect of another. Effector pairs with inverse roles exemplify the regulatory complexity employed by *L. pneumophila*.

SidM is one of the first *L. pneumophila* effectors to target Rab1 and is recruited to the LCV within 0 – 4 hours of infection (Müller et al., 2012). This protein binds PI4P on the LCV and

acts as a GEF toward Rab1, thereby targeting it to the pathogen vacuole. Once Rab1 has been activated on the LCV surface, SidM attaches an AMP moiety to Rab1 in a process called AMPylation (Hardiman and Roy, 2014) (Figure 2.4). It has been known for some time that AMPylated Rabs resist the activity of GTPase activating proteins (GAPs) and are thereby maintained in an activated state. Until recently, however, it was not known whether AMPylated Rab1 is maintained in an active conformation that can interact with GTP-state specific proteins. A recent study using molecular dynamics and umbrella sampling has shown that the AMP moiety promotes an active-like conformation of Rab1 regardless of the identity of bound nucleotide. This occurs by a stacking interaction between the adenine base of AMP and Phe45 on Rab1, which rigidifies the switch II region of Rab1 (Figure 2.5). Thus, SidM drives the localization of Rab1 to the LCV and maintains it in an active conformation (Luitz et al., 2016).

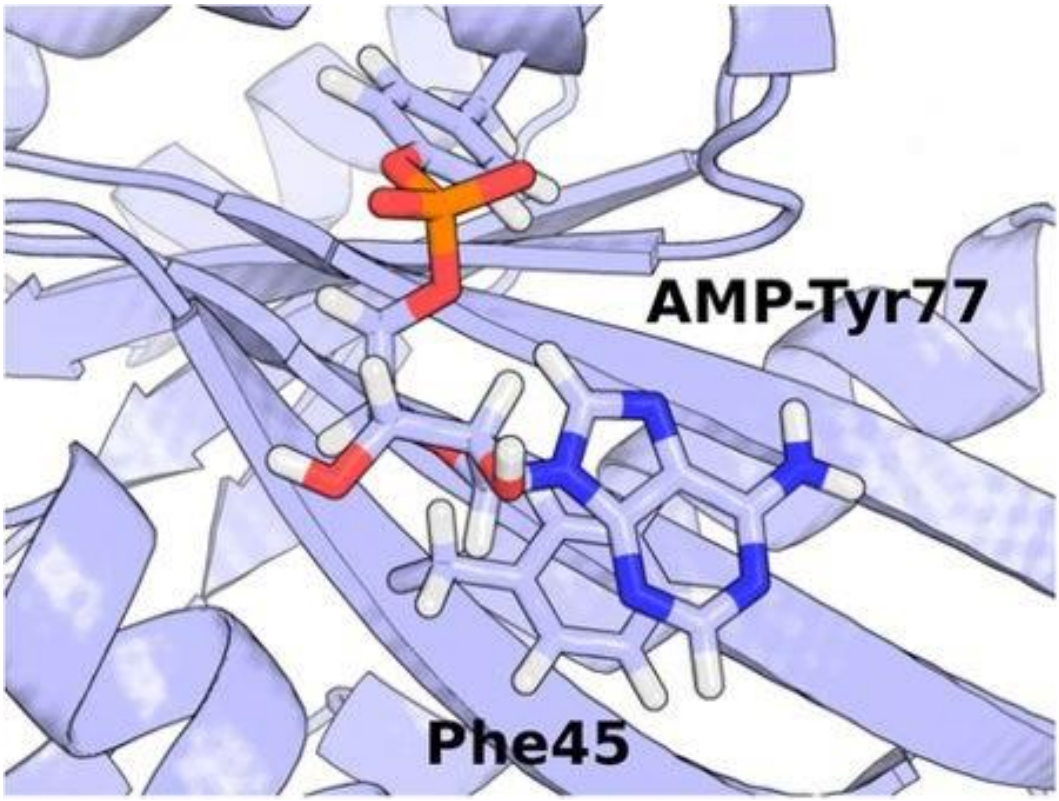

Figure 2. 5 AMPylated Rab1 is retained in an active conformation by a stacking interaction between the adenine base of AMP and the sidechain of Phe45. Figure adapted from (Luitz et al., 2016).

Activated Rab1 promotes tethering of ER-derived vesicles to the LCV, which then fuse in a process involving the noncanonical pairing of Sec22b with Stx3 syntaxin proteins (Arasaki et al., 2012). Tethering of ER-derived vesicles to the LCV depends on the exocyst, a protein complex that drives fusion of exocytic vesicles with the plasma membrane. Specifically, Sec5, Sec6 and Sec15 form a complex with Rab1-SidM-Stx3 and recruit Sec22b-positive vesicles to the LCV. The fact that only a subset of exocyst subunits are implicated in this fusion process may point to a special role for these proteins in membrane tethering. Since Sec5, Sec6 and Sec15 all bind SidM, these proteins may function independently in the recruitment of ER-derived vesicles to the LCV (Arasaki et al., 2018).

Since Rab1 AMPylation induces an active conformation regardless of the bound nucleotide, one may question the importance of GEF activity in directing Rab1 to the LCV. Indeed, studies separating GEF- from AMPylating activity in SidM have shown AMPylation to be critical for Rab1 localization to the LCV (Hardiman and Roy, 2014). This suggests that AMPylation of Rab1 can occur without activation, and that this modification reduces the affinity of GAP for Rab1 to impede membrane extraction by GDI.

Despite hydrolyzing bound GTP to GDP, cellular GAPs acting on AMPylated Rab1 are unable to reverse the effect of SidM (Luitz et al., 2016). Consequently, removal of AMP becomes a new prerequisite to deactivation of Rab1 in infected cells. It is intriguing to note that *L. pneumophila* also encodes an effector responsible for AMP removal. This effector is called SidD and is secreted into the host cell at a later stage than SidM (Figure 2.4). SidD is a de-AMPylase with a catalytic domain like that of metal-dependent protein phosphatases (PPMs). Bacterial PPMs have three metal binding sites, designated M1 – 3, with a conserved aspartate that coordinates M1 and M3. This residue was found to be absent in the crystal structure of SidD, shifting the M1 site and abolishing metal binding at M3 entirely (Figure 2.6) (Chen et al., 2013). Thus, SidD has been called a modified PPM that has adapted the ability to recognize Rab1.

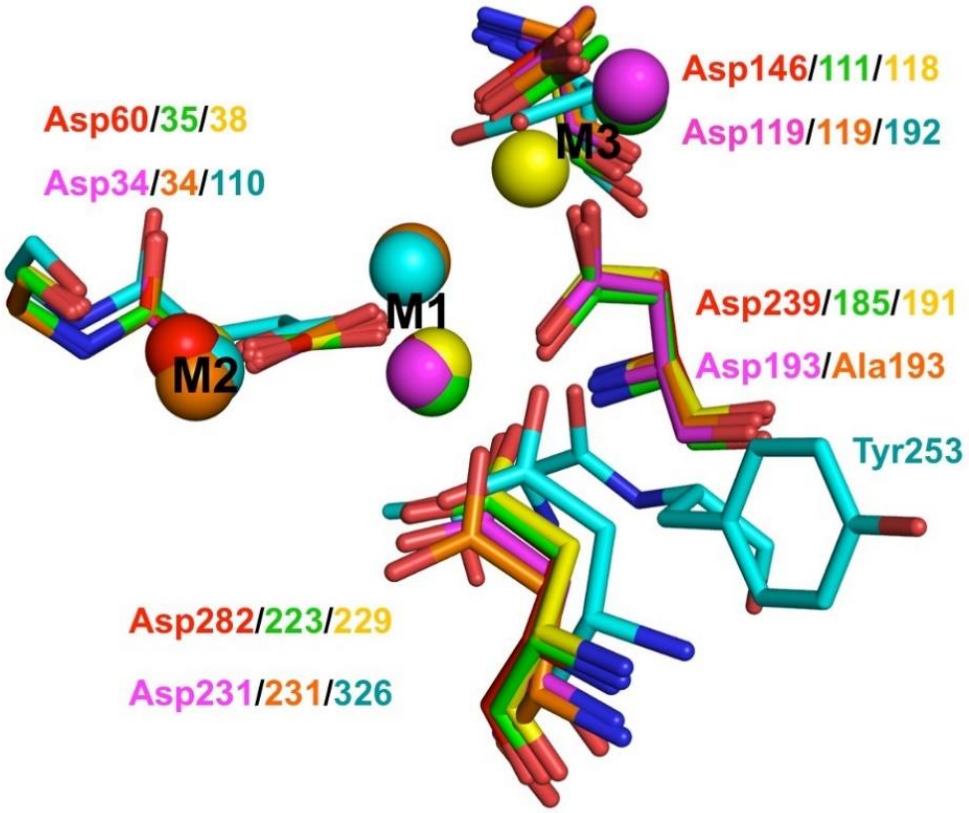

Figure 2. 6 Structural overlay of the catalytic sites in several metal-dependent protein phosphatases (PPMs) with coordinating residues depicted as sticks. SidD (cyan) is structurally homologous to the tPhpA(D193A) mutant. Neither model has a residue to coordinate M1 and M3, leading to a similar shift in the M1 ion and the absence of an M3 ion for both. The remaining structures (yellow, green, magenta and red) contain the M1 and M3 coordinating aspartate residue. Figure adapted from (Chen et al., 2013).

L. pneumophila also encodes an effector called LepB that antagonizes the GEF activity of SidM on Rab1 by acting as a GAP (Figure 2.4). The catalytic mechanism of LepB depends in part on its Arg444 sidechain, which coordinates the α- and β-phosphates of GTP. Meanwhile, Glu449 of LepB orients Gln70 of Rab1 which, in turn, orients a water molecule for attack on the GTP γ-phosphate. Surprisingly, Gln70 sits within a DXXGQ motif that is typically used by host GAPs for binding Rabs. Thus, a feature that distinguishes LepB from host GAPs is its use of Gln70 in catalysis rather than binding (Figure 2.7) (Mishra et al., 2013). The presence of effector proteins that reverse both the AMPylating and GEF activity of Rab1 highlights the significance of Rab1 regulation during infection.

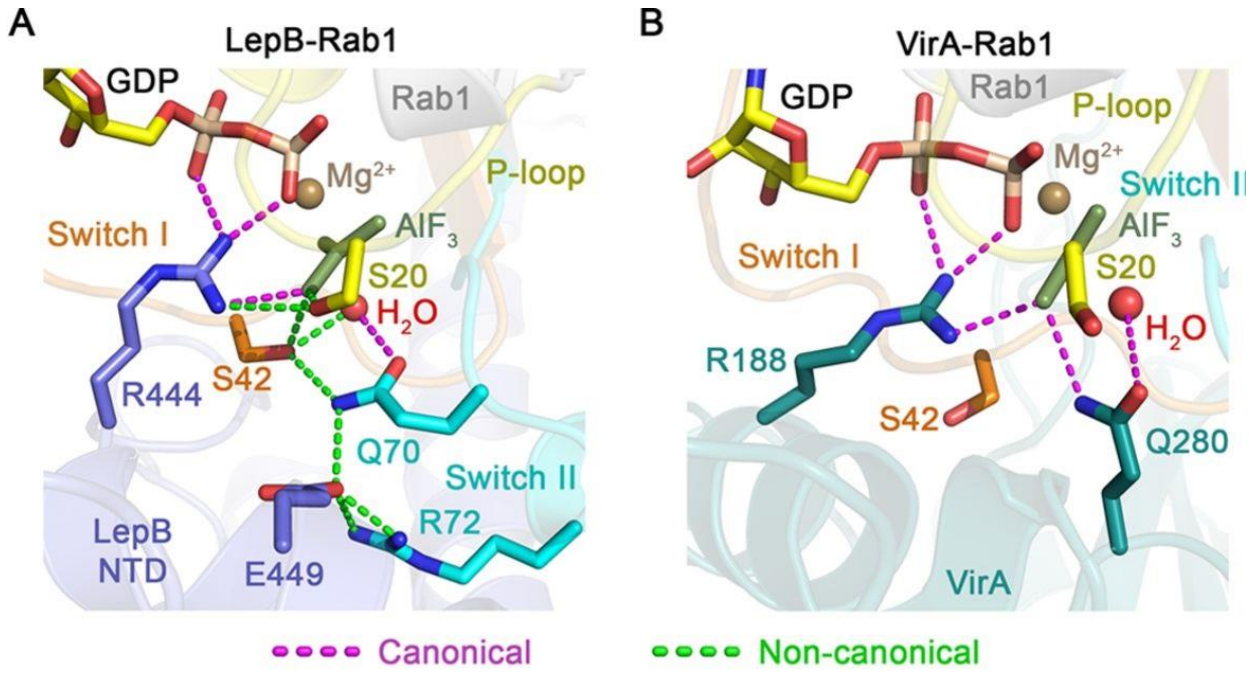

Figure 2. 7 Structural comparison of (A) the LepB-Rab1 and (B) VirA Rab1 active sites. Note that Gln70 of Rab1 and Gln280 of VirA coordinate an ordered water molecule in a similar fashion. Canonical and non-canonical interactions are shown in magenta and green, respectively. Figure adapted from (Mishra et al., 2013).

Another *L. pneumophila* effector that covalently modifies Rab1 is AnkX. This protein retains Rab1 in an inactivated state by phosphorylcholination (PCylation) of a serine (Ser79) residue adjacent to the tyrosine (Tyr80) residue to be AMPylated by SidM. Modification of these residues appears to be mutually exclusive (Neunuebel et al., 2012). That is, PCylated Rab1 cannot be AMPylated and vice versa. The proximity of these residues likely prevents such modifications from taking place simultaneously due to steric hindrance (Neunuebel et al., 2012). AnkX has a slight preference for GDP-loaded Rab1 over its active counterpart and, importantly, prevents the interaction of Rab1 with GDI (Goody et al., 2017; Mukherjee et al., 2011). Therefore, PCylation prevents inactive Rab1 from membrane extraction. This modification has also been found to prevent the interaction of LepB with Rab1 (Hardiman and Roy, 2014). Considering the preference of AnkX for GDP-Rab1 and the role of LepB in hydrolyzing bound GTP to GDP, it is interesting that GTP hydrolysis would be inhibited by a modification occurring predominantly on GDP-Rab1. Apparently, AnkX retains inactivated

Rab1 on the LCV and prevents its removal by LepB. Although the purpose of deactivating and immobilizing Rab1 is unclear, it would render this GTPase inaccessible to the host.

PCylation of Rab1 by AnkX is reversed by the *L. pneumophila* effector, Lem3, which acts as a dephosphorylcholinase (dePCylase). The *lem3* gene is separated from *ankX* by a 7bp intergenic region and these genes are transcribed convergently (Tan et al., 2011). Neither AnkX nor Lem3 are required for the intracellular replication of *L. pneumophila* in *D. discoideum* or bone-marrow derived macrophages (Tan et al., 2011).

LidA is an *L. pneumophila* effector that localizes to the LCV early during infection by an interaction with PI3P or PI4P. LidA recruits GDI-free Rab1 to the LCV, bringing Rab1 and SidM into proximity which, in turn, leads to the activation and AMPylation of LCV-associated Rab1. The crystal structure of LidA reveals the presence of two twisted β-hairpin repeats (TBR), one of which is involved in binding Rab1. This interaction orients the central domain of LidA to position Rab1 for membrane anchoring (Meng et al., 2013). Thus, although SidM retains Rab1 on the LCV, it is LidA that drives its recruitment there.

Another small GTPase implicated in the secretory pathway is Arf1. The *L. pneumophila* effector RalF retains Arf1 on the LCV using GEF activity, in much the same way that SidM recruits Rab1 (Nagai et al., 2002). Activated Arf1 facilitates secretion of vesicles from the LCV to the ER. Thus, Rab1 and Arf1 act in concert to promote a dynamic interplay of ER-LCV traffic and expand the pathogen vacuole (Kagan and Roy, 2002; Mizuno-Yamasaki et al., 2012). A recent study on small molecule inhibitors of ArfGEFs showed the GEF domain of RalF to resist inhibition by Brefeldin A. This result contrasted with other ArfGEFs, for which membrane binding had little impact on the efficacy of small molecule inhibitors. Full-length RalF harbors a capping domain with aromatic residues that mimic the Arf1 binding interface. Inhibition of RalF by Brefeldin A appears to depend on this capping regulatory domain (Benabdi et al., 2017).

2.1.7. Hijacking the host ubiquitination machinery

Ubiquitination is one of the best studied post-translational modifications of proteins. In eukaryotes, ubiquitin moieties are attached to target proteins in a complex process involving

E1-activating, E2-conjugating and E3-ubiquitin ligase enzymes. Since ubiquitination is important in the regulation of nearly all eukaryotic pathways, bacterial pathogens have evolved strategies to disrupt this process and use it for their benefit. *L. pneumophila* predominantly uses effector proteins that act as E3-ligase mimics or deubiquitinase (DUB) enzymes to interfere with host ubiquitination. Although the role of *L. pneumophila* effectors in ubiquitination has recently been reviewed in detail (Qiu and Luo, 2017b), some new findings have emerged in the last year.

L. pneumophila uses effectors of the cycle inhibiting factor (Cif) family, which deamidate a conserved glutamine residue on ubiquitin. This behaviour contrasts that of typical Cif family effectors, where NEDD8 harbors the glutamine to be modified (Valleau et al., 2018). Until this point, Cifs were found exclusively in pathogenic bacteria encoding a type III secretion system. Two *L. pneumophila* Cifs, MavC (*lpg2147*) and MvcA (*lpg2148*), have been structurally characterized by X-ray crystallography. These proteins show remarkable specificity for ubiquitin and exhibit no activity against NEDD8. To determine whether the ability to distinguish ubiquitin from NEDD8 could be traced to the binding mode, NMR spectroscopy was used in conjunction with prior knowledge of the Cif:ubiquitin complex. No major differences were observed for the MavC/MvcA:ubiquitin spectra relative to Cif:ubiquitin, attributing specificity to the local architecture of the active site rather than a different overall binding strategy (Valleau et al., 2018). This notion has since been challenged in a study using ITC to demonstrate that the C-terminal tail of ubiquitin allows MvcA to distinguish it from NEDD8 (Zhu et al., 2018). The noncanonical deamidation of ubiquitin by MavC was found to increase the rate of Ube2N ubiquitination and was therefore thought to increase the reactivity of Ube2N-conjugated ubiquitin, perhaps even in the absence of native target (Valleau et al., 2018). Furthermore, MavC activity has been found to dampen NF-κB signaling and prevent the formation of K63-linked ubiquitin chains on proteins belonging to pathways derived from NF-κB signaling. Finally, a third gene adjacent to MavC and MvcA, *lpg2149*, encodes an effector that inhibits the deamidase activity of these *L. pneumophila* Cifs (Valleau et al., 2018).

Legionella also produces dedicated deubiquitinases (DUBs) to remove ubiquitin from target proteins. One such effector, LotA, is comprised of two ovarian tumour (OTU)-like domains harboring catalytic cysteines (C13 and C303). Proteins belonging to the OTU

superfamily are cysteine proteases with DUB activity. The C-terminal catalytic cysteine of LotA, C303, acts in conjunction with a PI3P-binding segment to guarantee removal of ubiquitin from the LCV. In the N-terminal domain, C13 cleaves K6-linked ubiquitin chains. LotA appears to function synergistically with SidEs, because inactivation of both proteins in *L. pneumophila* causes a greater defect in bacterial replication than either protein individually (Kubori et al., 2018). Since LotA disassembles K-linked ubiquitin chains, this enzyme is unlikely to show activity against the S-linked ubiquitin generated by the SidE-family of effectors (see section 2.2.2.3). Thus, LotA is more likely to act during early stages of infection, before ubiquitination of substrates by SidEs has occurred. This distinction of LotA DUB activity from that of SidE-family effectors can be justified by noting that LotA must bind PI3P to remove ubiquitin from the LCV. Since PI3P is enriched on the early LCV, LotA likely carries out its DUB activity at early stages of bacterial vacuole formation (Kubori et al., 2018).

Another DUB encoded by *L. pneumophila*, SidJ, is unique in its ability to cleave PR-Ub linkages and shares a preference with SidE DUBs for K63-linked polyubiquitin chains. Although the catalytic mechanism of SidJ remains unclear, this enzyme does not act as a thiol- or metalloprotease and may therefore utilize a novel mechanism for cleavage of isopeptide/phosphoribosyl ubiquitin linkages. The DUB activity of SidJ is critical for *L. pneumophila* infection and reverses the ubiquitin-linkages generated by SidE-family enzymes (the N-terminal DUB activity of SidEs does not act on PR-Ub linkages) (Ronau and Hochstrasser, 2017). A detailed understanding of the SidJ DUB activity will provide insight into how *L. pneumophila* regulates its own pattern of ubiquitination.

2.1.8. Interference with endocytic and retrograde trafficking pathways

L. pneumophila disrupts secretory (anterograde), endocytic and retrograde trafficking pathways in macrophages. The secretory pathway defines transport of newly synthesized proteins or lipids from the ER to their cellular destinations (Qiu and Luo, 2017a). Endosomal trafficking describes the process in which cargos brought into the cell are transported to their designated organelles or destroyed within lysosomes (Allgood and Neunuebel, 2018). Finally, retrograde trafficking is the transport of endosomal cargo back to the ER (Elwell and Engel,

2018). Rab1 promotes (secretory) ER to Golgi traffic by recruiting the tethering factor p115 and v-SNAREs to ER-derived vesicles (Allan et al., 2000a). In the following section, an overview of the methods used by *Legionella* to hijack endocytic and retrograde trafficking is given.

Endocytic trafficking can be divided into three branches. In one, early endosomes are converted to late endosomes by a Rab-switching mechanism (Rink et al., 2005). Late endosomes then fuse with lysosomes, which harbor lytic enzymes to degrade the engulfed cargo. Another branch of the endocytic pathway is retrograde trafficking or the transport of early endosomes through the *trans*-Golgi network to the ER. Finally, early endosomes can be rerouted to the plasma membrane in a process called recycling endocytosis. These vesicular trafficking pathways are regulated by host Rab GTPases, which act as molecular switches and interact with effectors involved in membrane tethering and vesicle movement.

Recycling endocytosis occurs via fast and slow pathways. In the fast pathway, early endosomes are transported directly to the plasma membrane. The slow pathway moves early endosomes along membrane tubules to a perinuclear region called the endocytic recycling compartment (ERC) (Allgood and Neunuebel, 2018). *L. pneumophila* has been found to interfere with both the fast and slow pathways of recycling endocytosis, as evidenced by its effect on transferrin recycling. Transferrin is a substrate of recycling endocytosis and remains bound to its receptor on early endosomes until they fuse with the plasma membrane (Mellman, 1996). *L. pneumophila* diminishes the export of transferrin using the AnkX effector, which PCylates Rab35 and renders it inactive. Another Rab implicated in recycling endocytosis, Rab11a, was also found in association with the LCV and ubiquitinated in a manner dependent on the effector AnkB. Intriguingly, recruitment of Rab35 to the LCV depends on the presence of AnkB, which implies a role for ubiquitination in this process (Allgood and Neunuebel, 2018).

The LCV does not bind Rab5 or Rab7 and therefore deviates from the canonical endosome maturation process. In non-infected cells, Rab5 is recruited to early endosomes by the Rab5 GEF, Rabex-5, and promotes their fusion using an effector called Rabaptin-5 (Stenmark et al., 1995). Another Rab5 effector is the early endosome antigen 1 (EEA1), which has been found to play a role in tethering early endosomes to one another prior to fusion (Christoforidis et al.,

1999). Rab7 is later recruited to the maturing endosome by a Rab7 GEF called the Mon1-Ccz1 complex. The targeting of Mon1-Ccz1 to endosomes depends on Rab5-GTP and PI3P (Poteryaev et al., 2010). Thus, avoidance of Rab5 recruitment and activation prevents Rab7 accumulation on endosomes. *L. pneumophila* exploits this interdependency of endosomal Rabs using the VipD effector, which binds Rab5 and prevents its association with Rabaptin-5 and EEA1 (Ku et al., 2012). This may help the LCV avoid fusion with early endosomes while preventing steps characterizing the development of late endosomes or autophagosomes. Furthermore, Rab5-activated VipD uses phospholipase A activity to catalyse the removal of PI3P from endosomal membranes, thereby protecting the LCV from downstream fusion events (Lucas et al., 2014). Another effector, lpg0393, has also been found to infiltrate endocytic trafficking. Specifically, lpg0393 is structurally related to Rabex-5 and interacts with Rab5, Rab21 and Rab22. By exerting GEF activity on these Rabs, lpg0393 competes with Rabex-5 for their activation (Sohn et al., 2015).

Legionella also avoids autophagy using the secreted effector RavZ, which binds and liberates phosphatidylethanolamine (PE)-conjugated microtubule-associated protein 1A/1B light chain 3B (LC3) from the LCV. In non-infected cells, LC3 associates with the phagophore membrane using its PE tail. Lipidation of LC3 facilitates autophagosome biogenesis and the recruitment of autophagy cargo (Pankiv et al., 2007; Xie et al., 2008). Another autophagy-regulating protein, ATG4B, removes PE from LC3 in such a way that re-conjugation is possible (Nair et al., 2012; Nakatogawa et al., 2012). Although the purpose of PE deconjugation by ATG4B remains unclear (Agrotis et al., 2019), it is likely a form of regulatory control over autophagic trafficking. Unlike ATG4B, RavZ cleaves LC3 before the C-terminal glycine. This prevents LC3 from reattaching to PE and thereby reduces the pool of this autophagic component available to the host (Kwon and Song, 2018). Two structure-based hypotheses have been given for the mechanism of LC3 cleavage by RavZ. In one, RavZ is tethered to the membrane by two N-terminal and one C-terminal LIR motifs prior to catalysis. The other model suggests that a hydrophobic α-helix within RavZ separates LC3-PE from the membrane and exposes the cleavage site (Kwon and Song, 2018). A structure of RavZ in complex with LC3 is likely to clarify our understanding of this process.

It has been well established that retrograde trafficking impedes the survival of *L. pneumophila* within the host (Bärlocher et al., 2017). The polyphosphate-5-phosphatase, Dd5P4 (*D. discoideum*), and its human homologue, OCRL, play essential roles in retrograde transport and are recruited to the LCV at early stages of infection (Welin et al., 2018). An interaction of the LCV with these 5-phosphatases has been shown to restrict intracellular replication (Weber et al., 2009). A recent study using quantitative imaging flow cytometry has shown that Dd5P4 inhibits the recruitment of calnexin and Rab1 to the LCV but has no effect on Rab7/Rab8 accumulation. Since ER membranes are recruited to the LCV during maturation, calnexin can be considered a marker for the typical development of this vesicle. An absence of calnexin on the LCV therefore suggests that ER-interactions have been hampered. Such a disruption to LCV maturation is likely to affect replication, as this vacuole must develop into a niche that accommodates bacterial growth and division. Still, the mechanism relating calnexin recruitment to the LCV and a replication defect in *L. pneumophila* remains to be elucidated.

In addition to its role in preventing LCV maturation, Dd5P4 enhances localization of the sorting nexin, Vps5, to the LCV (Welin et al., 2018). This suggests that Dd5P4 acts prior to formation of the retrograde trafficking vesicle. Analysis of retromer components (Vps26, Vps29 and Vps35) present on the LCV in cells lacking Dd5P4 showed that Vps26 levels are increased while Vps35 is decreased. This differential regulation of retromer components by Dd5P4 may arise from distinct retrograde coat complexes, which harbor a unique complement of cargo recognition subunits (Welin et al., 2018). That is, LCV formation may depend on the recruitment of a specific retrograde coat complex. The actin nucleator, WASH, was also found to restrict intracellular replication of *L. pneumophila* (Welin et al., 2018). WASH drives the retrograde trafficking vesicle away from the endosomal network and its effect on replication indicates that even early stages of retrograde transport play a role in the survival of *L. pneumophila*.

Another *L. pneumophila* effector implicated in retrograde trafficking is RidL, which competes with TBC1d5 for binding to Vps29. Because TBC1d5 is a regulator of endosomal transport, RidL can be expected to block this retromer-mediated process (Yao et al., 2018).

2.2. X-ray crystallography in the study of host-pathogen interactions

2.2.1. Host-pathogen interactions

Intracellular bacterial pathogens engage the host machinery and disrupt autoimmune defenses. In many cases, the host-pathogen interaction occurs at the protein level. Bacterial proteins called effectors mimic, bind or modify various host proteins such as transcription factors, enzymes and regulatory elements. Functional analysis of effectors often requires identification of their host target(s). Once found, these target proteins implicate a pathway or process for further investigation. For example, an effector that interacts with a retromer component is likely to interfere with retrograde trafficking. Detailed structural data on the effector or effector-target complex can be used to determine how this interference occurs. An effector may adopt a fold known to carry out a specific function. Alternatively, an effector-target complex may show the target protein to be locked in an inactive conformation. By uncovering the consequences of target-binding, we can begin to infer the role of bacterial effectors in infection.

2.2.2. Applications of X-ray crystallography in the study of bacterial pathogenesis

X-ray crystallography has long been used as a tool to obtain structural data on effectors and the host-pathogen interface. This method allows the interaction of macromolecules to be visualized at atomic level resolution, provided such a complex can be captured in a crystal structure. Alternatively, ligand binding sites can be uncovered in models of independent effectors. Comparing the fold of an effector protein to that of a previously characterized model may give insights into its function. For example, if the unknown protein shares the fold of an enzyme, crystallography may be coupled with site-directed mutagenesis to reveal the mechanism. These strategies drive research forward by offering new lines of investigation into the protein of interest. Indeed, our appreciation of *L. pneumophila* pathogenesis has been largely aided by crystal structures. Many *L. pneumophila* effectors directly bind a eukaryotic protein to carry out their function in the host cell. Thus, crystal structures of effectors in

complex with their target often provide valuable insights into function. Herein, we review recent studies in which X-ray crystallography has advanced our understanding of *L. pneumophila* pathogenesis, with a special emphasis on host-pathogen interactions.

2.2.2.1. The secretory pathway

A key protein regulating secretory trafficking is the small GTPase, Rab1. In a non-infected cell, activated Rab1 localizes to the endoplasmic reticulum (ER) and Golgi apparatus (Plutner et al., 1991; Saraste et al., 1995), where it binds tethering factors (p115/v-SNAREs) to initiate the fusion of ER-derived vesicles with the Golgi (Allan et al., 2000b). *L. pneumophila* encodes several effectors to interfere with Rab1. One such effector is SidM/DrrA, which activates Rab1 by catalyzing the exchange of bound GDP for GTP while promoting the separation of Rab1 from GDI. Thus, by extracting GDP-Rab1 from GDI and catalyzing nucleotide exchange, SidM carries out both steps in the process of Rab1 activation. Structural studies on the SidM-Rab1 complex indicate that SidM displaces GDI from Rab1. A comparison of the SidM/Rab1 structure with a previously determined model of GDI/Ypt1:GDP revealed that Rab1 uses the same residues to bind both SidM and GDI. Further analysis led the authors to suggest that SidM evolved the ability to displace GDI by increasing its affinity for GDP-Rab1 (Suh et al., 2010).

Despite having GDF activity on GDP-Rab1, SidM binds GTP-Rab1 with much higher affinity. Therefore, while crystal structures of SidM/GTP-Rab1 have been obtained using relatively straightforward approaches, new methods of protein engineering were required to solve SidM in complex with GDP-Rab1. Specifically, bromoalkyl-lysine residues were incorporated at critical interfacial positions to generate covalent crystal contacts. This strategy led to the first structure of SidM:GDP-Rab1, which revealed that Mg^{2+} is essential for retaining Rab1 in the GDP-bound state and is coordinated by Ser22 and Asp63. The authors discovered that successive increases in the distances between these coordinating residues occur in structures of GppNHp-Rab1, SidM:GDP-Rab1 and SidM:Rab1 (Figure 2.8). This implies that dissociation of Mg^{2+} occurs prior to GDP removal, a critical new insight into the mechanism of nucleotide exchange by SidM (Cigler et al., 2017).

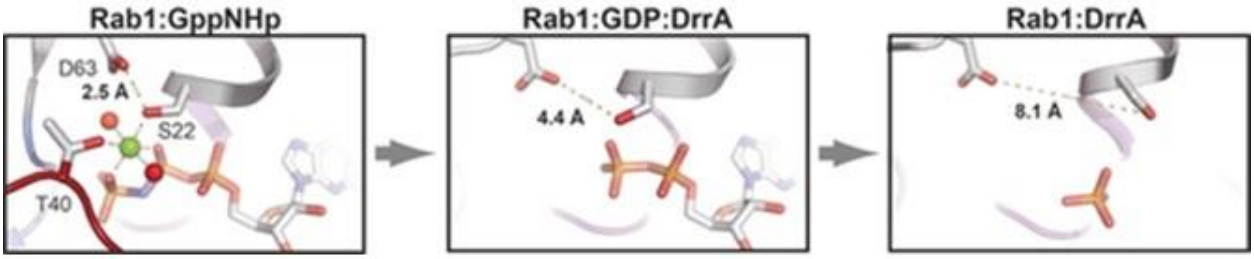

Figure 2. 8 Successive increases in the distance between Mg^{2+} coordinating residues (D63 and S22) in crystal structures of Rab1with a non-hydrolysable analogue of GTP analogue, GDP and no ligand. These findings suggest that Mg^{2+} exits the active site prior to GDP.

SidM not only activates but also modifies Rab1 by adding adenosine monophosphate (AMP) to a tyrosine sidechain located in the so-called "switch II" region, which undergoes significant conformational changes upon nucleotide exchange. AMPylation keeps Rab1 in an active conformation that does not depend on the identity of bound nucleotide. This modification also prevents activated Rab1 from interacting with the host protein, MICAL-3. Intriguingly, *L. pneumophila* encodes an effector called LidA that disrupts secretory trafficking by binding to active, inactive or even AMPylated Rab1. LidA has also been shown to bind Rab6 and Rab8, which are recruited to the LCV during infection (Schoebel et al., 2011). Eukaryotic Rab effectors typically use two α-helices to interact with a hydrophobic triad between the switch regions of the Rab protein. The crystal structure of LidA in complex with Rab8a shows an unusually large binding interface of four α-helices that extends beyond both switch regions and the hydrophobic triad of Rab8a (Figure 2.9). The presence of interacting residues outside the switch regions likely allows LidA to bind AMPylated Rab and leads to its picomolar affinity for GTP-Rab proteins. A consequence of this high binding affinity is limited dissociation of the complex, rendering Rab inaccessible to host GAPs (Schoebel et al., 2011). This may imply a role for LidA in recruiting Rabs to the LCV, although the exact function of LidA in *L. pneumophila* pathogenesis remains unclear. The crystallographic data obtained for the LidA/Rab8a complex reveals a unique binding mechanism that has garnered LidA the title of "Rab supereffector."

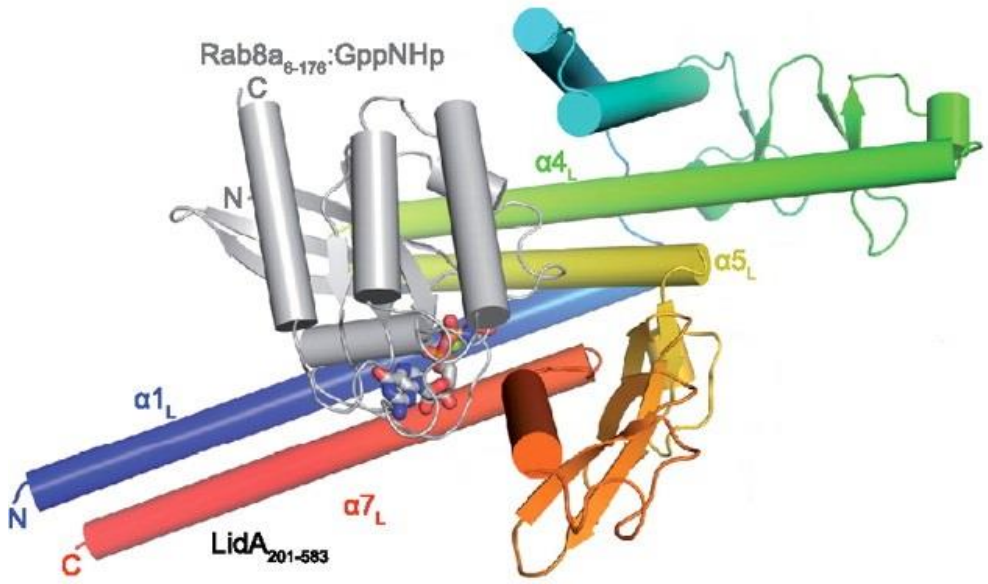

Figure 2. 9 Crystal structure of the LidA-Rab8a complex with Rab8a shown in grey and LidA in rainbow. The four interacting helices in LidA are labelled ($\alpha1_L$, $\alpha7_L$, $\alpha4_L$, $\alpha5_L$). Figure adapted from (Schoebel et al., 2011).

2.2.2.2. Effectors with protease activity

Several crystal structures of *L. pneumophila* effectors with protease activity have recently been determined. One of these structures was obtained in a study that coined the term "metaeffector" to describe effectors that mitigate the role of another. To fully characterize one of these effector/metaeffector pairs, the authors solved the crystal structure of LegL1 in complex with RavJ (Urbanus et al., 2016). Initially, the RavJ N- and C-terminal domains were solved independently. Structural homology between the N-terminus of RavJ and papain-like cysteine proteases (PLCP) allowed the authors to identify putative catalytic triad residues for this protein. Indeed, mutating any of these residues abrogated the growth defect in yeast cells expressing RavJ. Meanwhile, LegL1 is a leucine-rich repeat (LRR) protein that adopts a curved overall fold. In the LegL1/RavJ co-crystal, LegL1 arches over the RavJ catalytic triad, impeding its activity in a non-proteolytic mechanism of steric inhibition (Figure 2.10) (Urbanus et al., 2016).

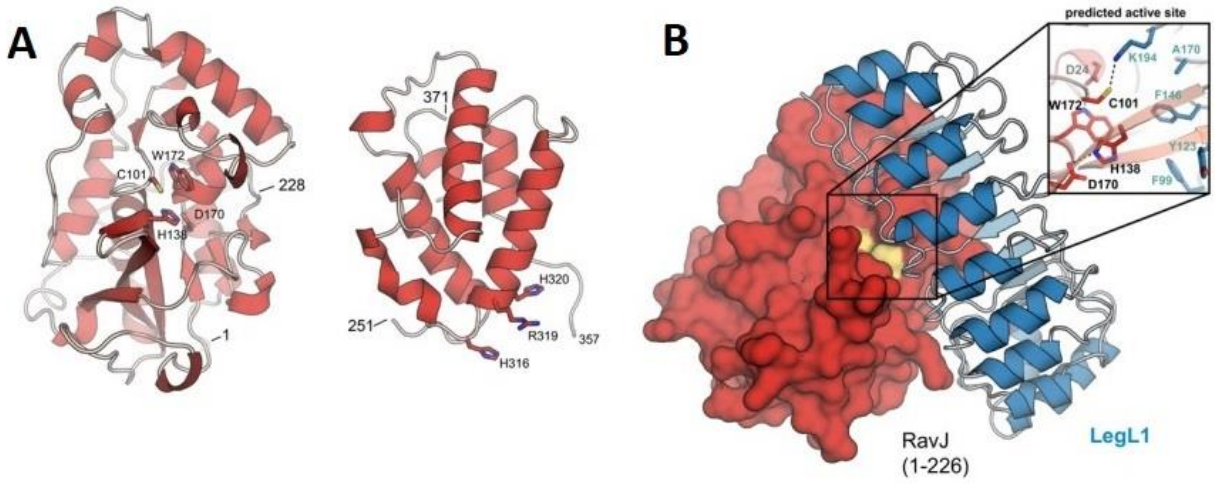

Figure 2. 10 Crystal structure of (A) RavJ and (B) the RavJ-LegL1 complex. Active site residues are shown as sticks. LegL1 arches over the active site of RavJ to prevent catalysis. Figure adapted from (Urbanus et al., 2016).

Lpg2622 is a substrate of the Lsp type II secretion system (T2SS) that also belongs to the PLCP family. Although the role of cysteine proteases in *L. pneumophila* pathogenesis remains unclear, a similar protein in *Streptococcus pyogenes* (SpeB) has been shown to degrade immunoglobulins and cleave cytokine precursors. These abilities help *S. pyogenes* evade host immunity and promote septic shock. Crystal structures have been solved for Lpg2622 in the apo-form and in complex with the E64 inhibitor. Lpg2622 shares the architecture of C1 family peptidases but contains a novel propeptide comprising a hairpin-turn-helix (β_1-β_2-α_3) fold. Importantly, the propeptide of Lpg2622 does not fully occupy the active site cleft, although it makes extensive interactions with its catalytic C-terminus. This contrasts with the propeptides of cathepsin-L-like or B-like enzymes, which extend deep into the catalytic pocket to inhibit enzyme activity. Furthermore, removing any secondary structure elements from the Lpg2622 propeptide leads to the formation of inclusion bodies. These findings suggest the propeptide is essential for proper folding of Lpg2622 and is unlikely to be cleaved. The Lpg2622-E64 complex structure corroborates this notion by being the first PLCP-inhibitor model with the propeptide intact. Importantly, the β_2-strand of the Lpg2622 propeptide plays a regulatory role in catalysis, as abolishing its interaction with the catalytic domain increased enzymatic activity

of Lpg2622 (Gong et al., 2018). These findings show a unique modification on PLCP family enzymes where the propeptide acts as a non-cleavable regulatory element.

Other substrates of the *L. pneumophila* T2SS are the predicted aminopeptidase and acyltransferase, LapA and PlaC, respectively. A recent study on these proteins showed that mutating LapA increases the levels of PlaC mRNA in *L. pneumophila* and vice versa, suggesting that these proteins compensate for one another during infection. Moreover, mutating both LapA and PlaC significantly limits the invasion of *Acanthamoeba castellanii* by *L. pneumophila*. LapA cleaves various aminopeptides including leucine, valine, isoleucine, aspartate, methionine and phenylalanine as a nutrient source for *L. pneumophila* but the role of this function in host cell invasion remains unclear. To establish the molecular basis for its broad substrate specificity, the authors solved the crystal structure of LapA and showed it to contain a deep hydrophobic cavity adjacent to the active site, which could accommodate aromatic and aliphatic side chains. Comparison of the LapA model with a previously solved structure of LapB shows a wider and more negatively charged pocket in the latter, consistent with its ability to cleave near lysine and arginine residues. This is a case where X-ray crystallography has been used to show how small structural differences between similar proteins can produce significant functional consequences. Although the mechanism of complementation between PlaC and LapA remains to be deduced, it is clear from this study that enzymes secreted by the T2SS play an important role in *L. pneumophila* invasion.

2.2.2.3. The ubiquitin pathway

Many *L. pneumophila* effectors have been found to subvert host ubiquitination. One of these is AnkB, an effector that promotes ubiquitination by interacting with the eukaryotic Skp1-Cullin-F box (SCF) ubiquitin ligase. AnkB facilitates the attachment of poly-ubiquitinated proteins to the LCV. These proteins are ultimately degraded into free amino acids, which serve as an energy source for *L. pneumophila*. The role of AnkB in pathogenesis is essential for *L. pneumophila* replication in human macrophages, making this protein a particularly interesting candidate for investigation.

Until recently, our understanding of how AnkB usurps the function of SCF remained unclear. A crystal structure of AnkB in complex with Skp1 revealed that N-terminal residues 2 – 165 of AnkB adopt an F-box fold that is nearly identical to that of Skp2. The interaction between AnkB and Skp1 relies on a hydrophobic patch present in helices 1 and 2 within the AnkB F-box. This patch causes AnkB to form inclusion bodies when purified in the absence of Skp1. A high resolution (1 Å) crystal structure of the AnkB ankyrin domain (res 54 – 168) shows a C-terminal pentapeptide (QEEKI) from one molecule binding the curved surface of a neighboring chain. Although this interaction is unlikely to be of biological significance to AnkB, it does highlight the promiscuity of protein-protein interactions mediated by ankyrin domains. Furthermore, mutations destabilizing the ankyrin fold of AnkB were found to disrupt the intracellular replication of *L. pneumophila*. Although other effectors harboring F-box domains are encoded by *L. pneumophila*, only AnkB contains an attached ankyrin domain. These structural data show that AnkB docks to the SCF using an N-terminal F-box domain while recruiting host proteins for ubiquitination with its ankyrin domain. Thus, AnkB provides two examples of molecular mimicry by independent domains that perform a concerted function (Wong et al., 2017).

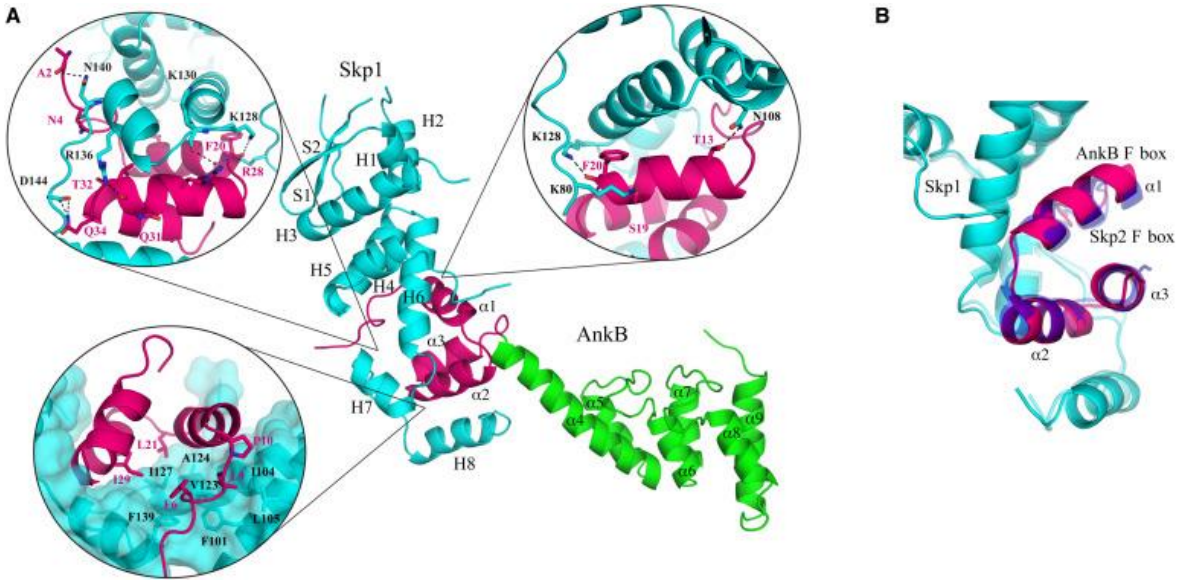

Figure 2. 11 Crystal structure of AnkB in complex with Skp1. (A) The ankyrin repeat region of AnkB is shown in green and the F-box is in pink. Skp1 is depicted in cyan. (B) Overlay of the Skp2 (dark blue) and AnkB (pink) F-boxes. Figure adapted from (Wong et al., 2017).

Another *Legionella* effector, RavN, is also involved in ubiquitination. Despite acting as an E3-ubiquitin ligase, RavN shares no primary sequence homology with any known E3. The crystal structure of RavN shows a similar fold to that seen in U-box E3 ligases, with the best alignment occurring at the central β-sheet region. This remote similarity between RavN and known E3s inspired the authors to search for additional effectors using hidden Markov model-based secondary structure analysis. Four new effectors involved in ubiquitination were discovered in this way (Lpg2370, Lpg2577, Lpg2498, and Lpg2452/SdcB). The number of effectors employed in this process highlights the significance of ubiquitination to the survival of *L. pneumophila*. Structural deviations from the classical fold may suggest that these effectors were obtained by horizontal gene transfer and structurally altered during their evolution. This work demonstrates how structural biology can be used to reveal the function of effectors showing only vague similarity to known E3s (Lin et al., 2018).

SidC is an *L. pneumophila* effector with E3 ubiquitin ligase activity that recruits ER-derived proteins to the LCV. A crystal structure of the SidC N-terminal ubiquitin ligase (SNL) domain exhibited no homology to known proteins. Multiple sequence alignment of SidC proteins from various *L. pneumophila* species revealed conserved catalytic residues of a cysteine protease. Structural analysis showed this putative catalytic triad (Cys46, His444, Asp446) to reside on a negatively charged surface at the center of the SNL domain. Surprisingly, SidC does not act as a protease but exhibits E3 ligase activity that depends on the catalytic cysteine (Cys46). The authors suggest that His444 and Asp446 increase the nucleophilicity of Cys46 and thereby facilitate the transfer of ubiquitin from E2 onto SidC (Hsu et al., 2014). This serves as an example of how structural biology can operate in concert with bioinformatics to reveal functionally important residues in a protein of interest.

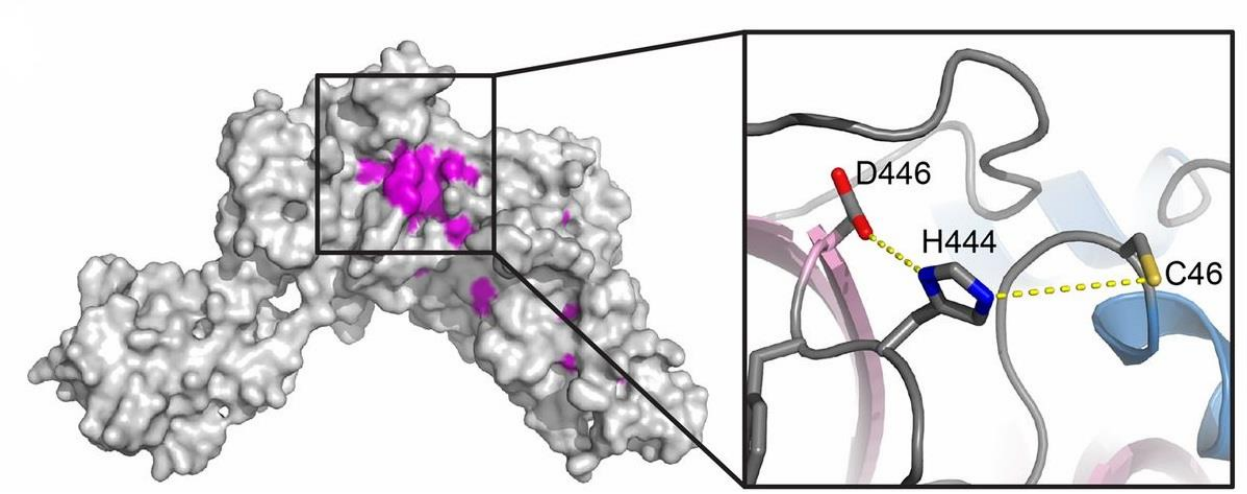

Figure 2. 12 Surface electrostatic rendering of the SidJ crystal structure showing a negatively charged surface (purple) containing putative catalytic triad residues. This area is enlarged on the right to show the catalytic residues and their hydrogen bonding pattern. Figure adapted from (Hsu et al., 2014).

Another interesting method of E3-ligase mimicry is demonstrated by the SidE family of effectors (SdeA, SdeB, SdeC and SidE), which ubiquitinate target proteins without the help of E1 or E2 enzymes. These effectors rely on mono-ADP-ribosyltransferase (mART) and phosphodiesterase (PDE) activities to carry out ubiquitination in an NAD^+-dependent manner. Specifically, mART activity catalyzes the formation of ADP-ribosylated ubiquitin (ADPR-Ub), which is cleaved into AMP and phosphoribosylated ubiquitin (PR-Ub) by the PDE. Finally, PR-Ub is attached to serine on a target protein (Akturk et al., 2018). The use of serine contrasts with the conventional approach, in which ubiquitin is linked to the target protein by a lysine residue.

Extensive structural studies have been carried out to explain the all-in-one approach to ubiquitination employed by SidEs. The crystal structure of SdeA contains a mono-ADP-ribosyltransferase (mART) domain that is divided into an α-helical lobe (AHL) and a mART core. The mART core is near the phosphodiesterase (PDE) domain of SdeA, while the AHL is distant from remaining regions of the protein (Figure 2.13) (Kalayil et al., 2018). Solution NMR structures showed the AHL to be adjacent to the mART core, however, suggesting that transient movements of this domain occur to facilitate NAD^+ binding and catalysis. Importantly, both mART and PDE domains were found to act independently of one another.

That is, a SdeA construct of mART only ubiquitinated target proteins if a construct of PDE was also present and vice versa. This suggests that ADP-ribosyltransferase and phosphodiesterase activities do not depend on the interaction of these individual domains.

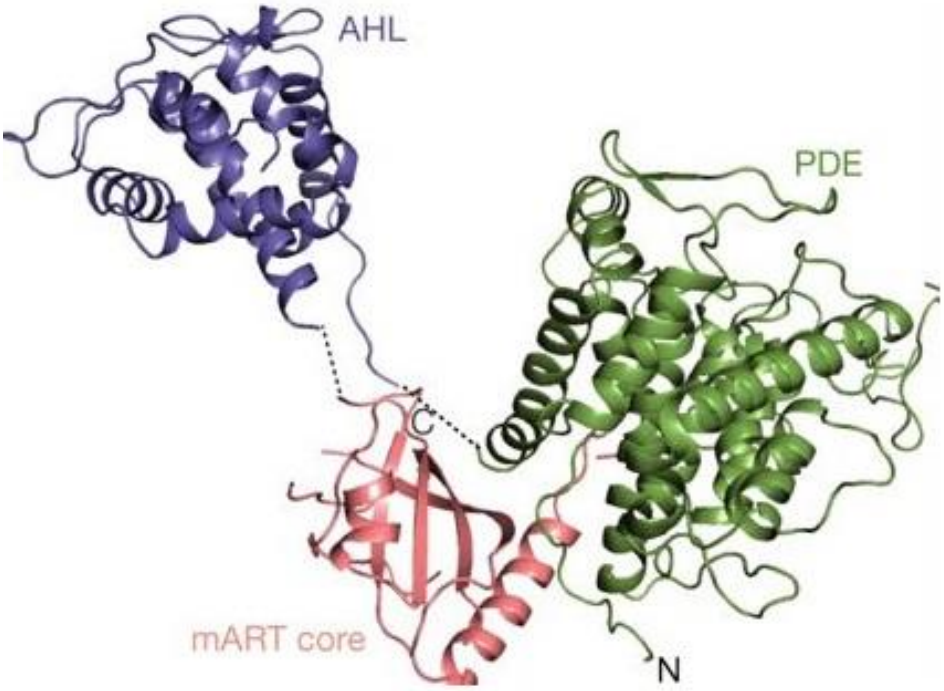

Figure 2. 13 Crystal structure of SdeA showing the mART core (pink) in proximity to the PDE domain (green), with the AHL (blue) distant from both. Solution NMR studies of this protein suggest the AHL can move near the mART core to facilitate catalysis. Figure adapted from (Kalayil et al., 2018).

A complex structure of SdeA-Ub-NADH provided new perspective on the ADP-ribosylation of ubiquitin at the mART domain. Although it had been accepted that Arg42 of ubiquitin attacks the N-ribose in an SN1 reaction, these atoms are too far apart in the complex to justify such a mechanism. To address this issue, molecular dynamics simulations were carried out using the SdeA-Ub-NADH complex structure with NADH replaced by NAD+. These studies revealed a nearby arginine residue, Arg72, which recognizes the strained nicotinamide moiety of NADH. Separation of nicotinamide from NAD+ destabilizes the interaction of Arg72 and allows Arg42 to move in the direction of ADP-ribose, bringing it into proximity with the N-ribose and thereby facilitating SN1 mediated catalysis (Dong et al., 2018).

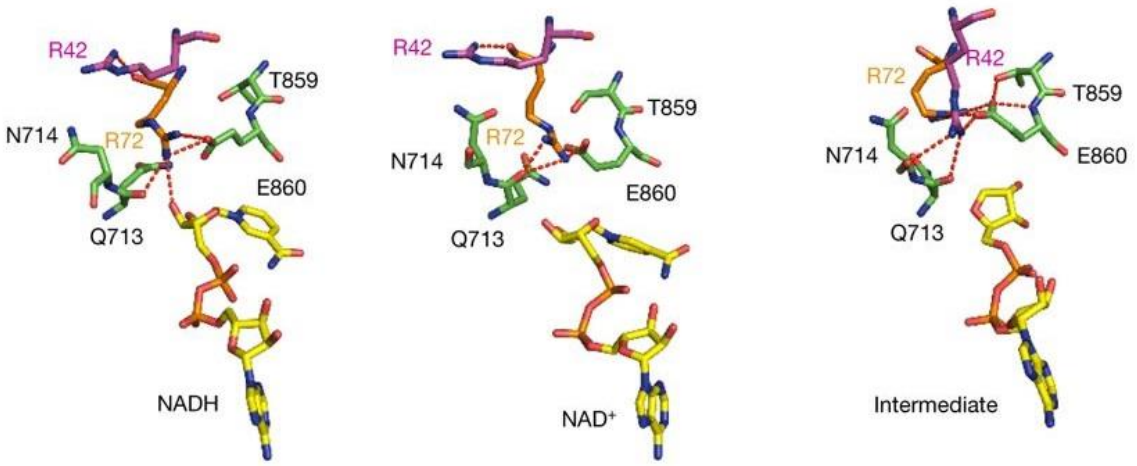

Figure 2. 14 Conformational changes of ubiquitin residues prior to ADP-ribosylation by SdeA. Arg72 initially recognizes the nicotinamide moiety of NADH (left) and moves away from the ligand once cleavage has occurred (middle and right). Meanwhile, Arg42 orients itself toward N-ribose after nicotinamide has been cleaved. Figure adapted from (Dong et al., 2018).

The crystal structure of inactive SdeD in complex with ADP-ribosylated ubiquitin provides insight into the interaction of ADP-ribose with the PDE domain. ADP-ribose contacts one glutamate and two histidine residues in the active site. These residues were invariant between SidE-family proteins, and the corresponding residues in SdeA (His277, His407 and Glu340) abolished phosphodiesterase activity when mutated to alanine. These data allowed the authors to propose a catalytic mechanism for the PDE domain, wherein Glu340 positions His277 for nucleophilic attack on the β-phosphate of ADP-ribose. This leads to the formation of a transient phosphoramidate intermediate, which is cleaved by a serine residue on the target protein (Akturk et al., 2018).

2.2.2.4. Endocytic, autophagic and retrograde trafficking pathways

L. pneumophila also disrupts endocytic and retrograde trafficking pathways in macrophages. Endosomal trafficking describes the process in which cargos brought into the cell are transported to their designated organelles or destroyed within lysosomes (Allgood and Neunuebel, 2018). Autophagy is a branch of endocytic trafficking that is dedicated to the

digestion of engulfed material. Conversely, retrograde trafficking is the transport of endosomal cargo back to the ER (Elwell and Engel, 2018). Crystal structures have added extensive detail to our understanding of how *L. pneumophila* uses effector proteins to disrupt these processes.

VipD is an *L. pneumophila* effector that binds Rab5 on endosomes and interferes with the Rab5-Rab7 conversion that characterizes endosomal maturation. Furthermore, Rab5-activated VipD uses phospholipase A (PLA_1) activity to catalyse the removal of PI3P from endosomal membranes and thereby protect the LCV from downstream fusion events (Lucas et al., 2014). The PLA_1 activity of VipD depends on its interaction with Rab5. A crystal structure showed VipD to harbor a surface loop, called the 'lid,' that occludes the active site. The lid is a common regulatory feature among phospholipases that undergoes conformational changes to expose the catalytic site in activated enzymes. It was suggested that the VipD lid changes its orientation upon Rab5-binding. A crystal structure of the Rab5/VipD complex was solved to determine the conformational effects of the Rab5 interaction. As expected, Rab5-binding repositions the VipD lid to expose its catalytic pocket. The interaction of Phe442 in VipD within a hydrophobic pocket on Rab5 leads to a cascade of conformational changes that ultimately displaces the β10-α14 (lid) loop, which would otherwise block the active site. These conformational changes were determined based on comparisons of the complex with a model of isolated VipD (Lucas et al., 2014). This study serves as an elegant example of how the information in different crystal structures can be combined to explain molecular rearrangements.

RidL is an *L. pneumophila* effector known to impede retrograde trafficking by engaging retromer components. The mechanism of inhibition has been revealed by two crystal structures. One model shows the RidL/Vps29-Vps35 complex (Romano-Moreno et al., 2017) and the other contains only RidL/Vps29 (Yao et al., 2018). Both structures provide insight into the host-pathogen interface by showing that RidL competes with the TBC1d5 regulator for binding to Vps29. This is made apparent by the similar conformations of binding loops present in RidL and TBC1d5. Since the retromer requires TBC1d5 for endosomal transport, RidL likely blocks retromer-mediated traffic by preventing the association of this regulator with Vps29 (Yao et al., 2018). These studies provide yet another example of molecular mimicry in effector-host interactions as revealed by X-ray crystallography.

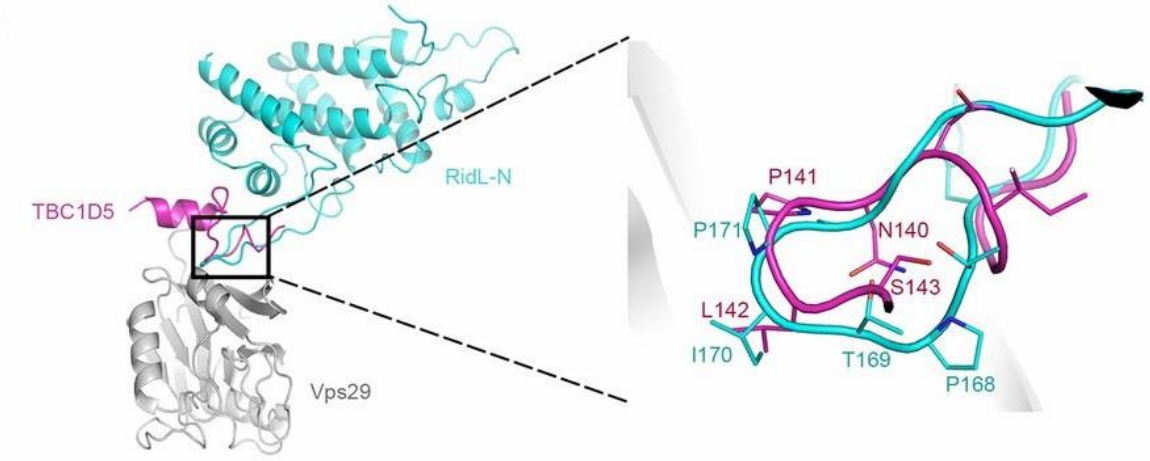

Figure 2. 15 Crystal structure of the RidL N-terminal domain (cyan) in complex with Vps29 (grey). The binding loop of TBC1d5 (magenta) is overlaid and shown in greater detail on the right. Figure adapted from (Yao et al., 2018).

Acidification of late endosomes and lysosomes depends on a membrane protein complex called the vacuolar H^+-ATPase (v-ATPase), which pumps protons into these compartments at the expense of ATP hydrolysis (Harrison and Muench, 2018). Lowering the pH of phagosomes is an essential part of autophagy, as lysosomal digestive enzymes have optimal activity under acidic conditions and production of reactive oxygen species (ROS) requires protons (Vieira et al., 2002). The v-ATPase is comprised of a membrane-embedded (V_0) and cytosolic (V_1) region consisting of three catalytic A/B heterodimers. These heterodimers each contain a nucleotide binding site that undergoes conformational changes to facilitate ATP hydrolysis by a rotary catalytic mechanism (Walker, 1998). *L. pneumophila* relies on the effector SidK to prevent acidification of the LCV. The mechanism by which SidK inhibits v-ATPase activity was revealed using single particle electron cryomicroscopy (cryo-EM) in conjunction with crystallography. A crystal structure of the SidK N-terminus reveals three α-helical bundles with two short helices in bundle I that deviate from the helical axis. The cryo-EM model of SidK/v-ATPase shows three SidK molecules bound to the catalytic A-subunits of the complex. This interaction is mediated by SidK residues in the short and long helices of bundle I. The C-terminus of SidK was poorly resolved in the cryo-EM map, suggesting flexible tethering of this molecule to the v-ATPase. Surprisingly, the SidK/v-ATPase interaction did not show any

significant conformational changes to v-ATPase. The authors found that SidK limited the flexibility in A/B heterodimers to lower the affinity of ATP for this region of the protein. This led to a ~40% decrease in v-ATPase activity, impeding acidification of the LCV without killing the host cell (Zhao et al., 2017).

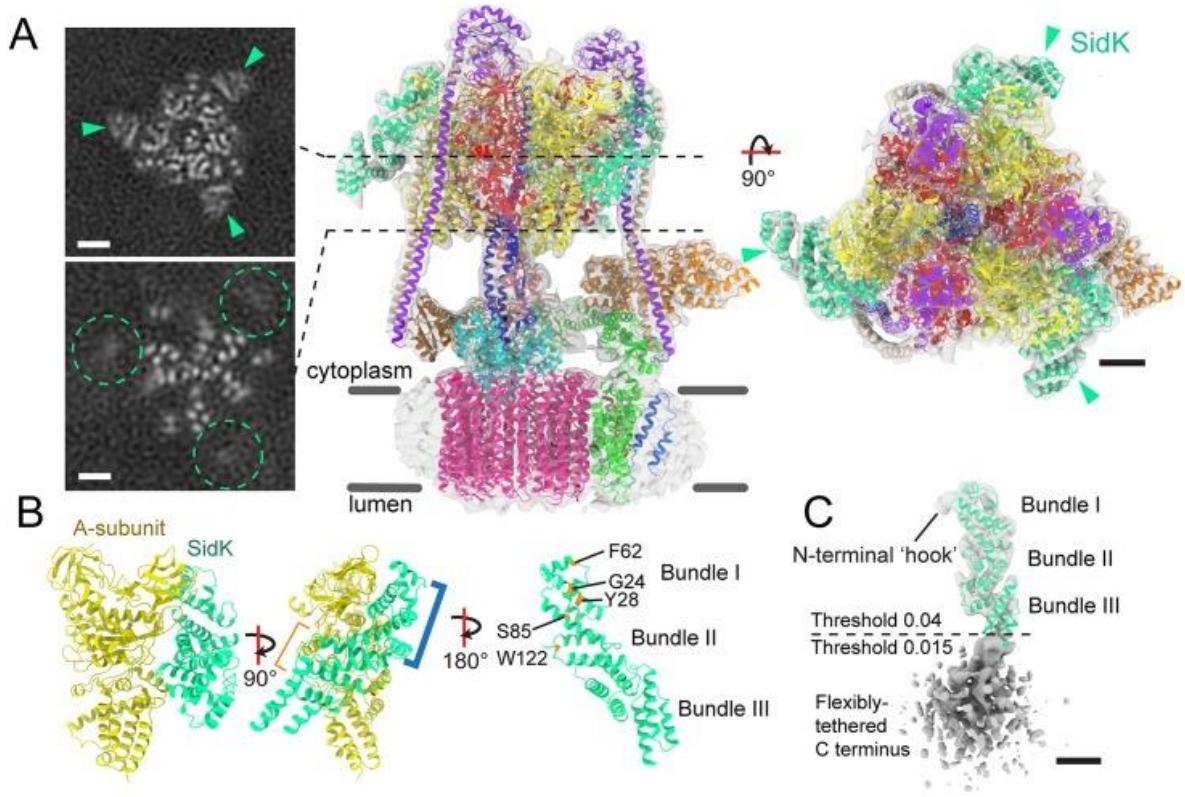

Figure 2. 16 Cryo-EM structure of SidK in complex with the vATPase. (A) Electron micrographs showing three molecules of SidK bound to the catalytic V1 region of the v-ATPase. SidK is well resolved at the N-terminus, where it contacts v-ATPase (teal arrowheads) but poorly resolved at the C-terminus (teal dashed circles). (B) Molecular contacts between SidK (teal) and the A-subunit of v-ATPase (yellow). Critical interacting residues within bundle I are denoted. (C) Cryo-EM map density showing a lack of definition at the C-terminus. Figure adapted from (Zhao et al., 2017).

RavZ in an *L. pneumophila* effector that irreversibly removes PE from the autophagy marker, LC3 (see section 2.1.8). To address whether RavZ depends on membranes for its activity, researchers generated liposome-free LC3-PE by a semisynthetic approach and exposed these compounds to RavZ. These studies showed that RavZ does not rely on

membranes to deconjugate PE from LC3-PE, leading the authors to suggest that RavZ extracts LC3-PE from membranes prior to initiating cleavage. The crystal structure of RavZ in complex with LC3-PE revealed an LC3-interacting region (LIR) motif and lipid-binding site (LBS) on RavZ, which appear to position the LC3 C-terminus for cleavage. The absence of a LBS in ATG4B may explain the different mechanism of these proteins. A hydrophobic α-helix within RavZ is thought to separate LC3-PE from the membrane and expose the cleavage site (Yang et al., 2017).

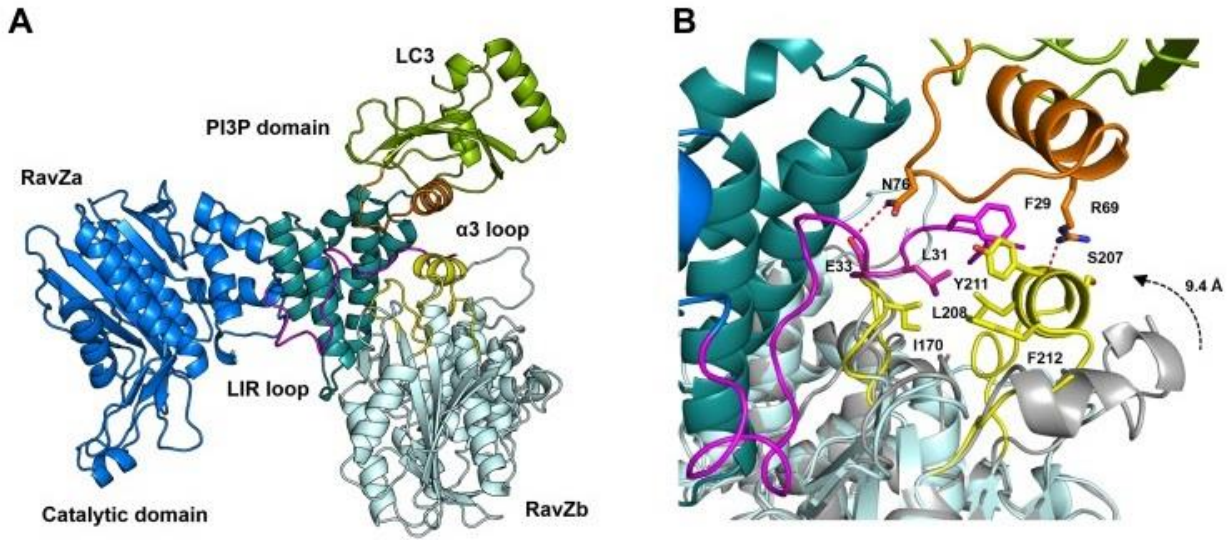

Figure 2. 17 Crystal structure of RavZ in complex with LC3. (A) RavZa contacts the α2-β3 loop (orange) of LC3 using its LIR loop (purple). The catalytic and PI3P-binding domains of RavZa are shown in blue and teal, respectively. LC3 is shown in green. (B) A symmetry related RavZb (cyan) contacts the LIR loop of RavZa with its α3 and β6-β7 loops (yellow). The flexible α3 could move to expose the LBS of RavZ. Figure adapted from (Yang et al., 2017).

2.2.2.5. Effector kinases and phosphatases

LegK4 belongs to the 'eukaryotic-like' family of kinases encoded by *L. pneumophila*. It is comprised of an N-terminal domain, a eukaryotic-like kinase domain (consisting of N- and C-lobes) and a C-terminal domain. A helical cap makes extensive contacts with and likely stabilizes the N-lobe of the kinase domain in LegK4. In contrast to conventional kinases, LegK4 does not depend on autophosphorylation of the activation loop to stabilize this segment of the protein. An extensive hydrogen bond network allows LegK4 to adopt an active

conformation in the absence of autophosphorylation. Furthermore, the LegK4 structure reveals a noncanonical dimeric interface that stabilizes the activation loop to facilitate catalysis (Figure 2.18). These structural observations help to explain the constitutive activity of LegK4 and reveal how fascinatingly an effector can evolve to suit the pathogen (Flayhan et al., 2015).

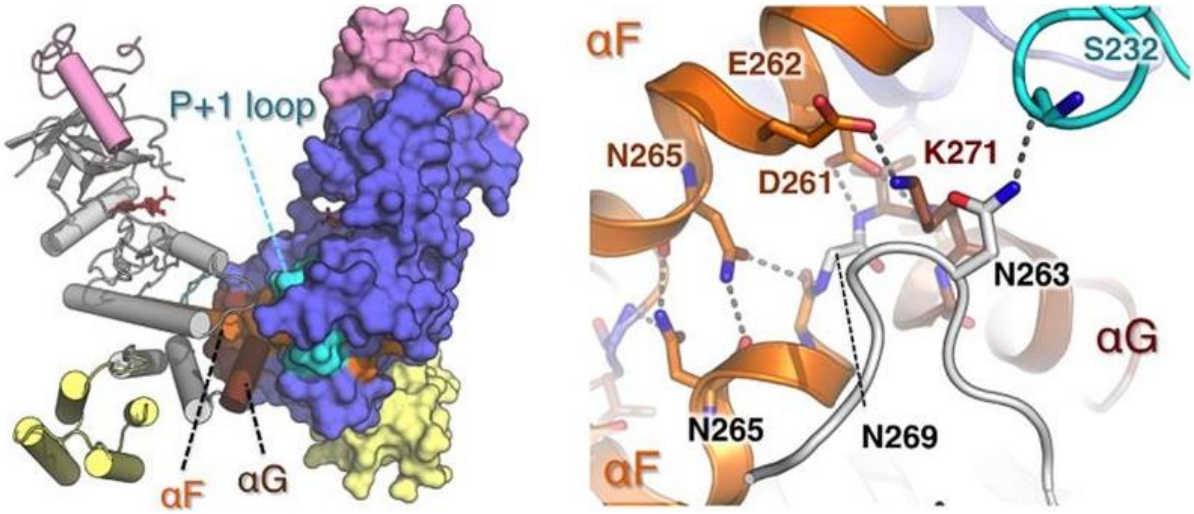

Figure 2. 18 Crystal structure of the LegK4 dimer with chain A shown as surface and chain B shown as ribbon (left). Interactions between αF helices (orange) at the dimer interface facilitate a hydrogen bond between Asn263 and Ser232 on the P+1 loop. This interaction is present in both apo- and substrate-bound forms, suggesting that dimerization has a stabilizing effect on the activation loop. Figure adapted from (Flayhan et al., 2015).

Ceg4 is a haloacid dehalogenase (HAD)-like phosphotyrosine phosphatase with activity against eukaryotic MAP kinases. The crystal structure of Ceg4 reveals an N-terminal HAD-like domain and a C-terminal domain with two transmembrane helices. Ceg4 appears to rely on its C-terminal domain for interacting with substrate, as deletion of this region abrogates binding to MAP kinases. Importantly, an α-helical cap domain that forms a lid over the active site was also observed in Ceg4. Although such caps have been observed in other HAD-like proteins, the Ceg4 cap takes on a novel fold that is compatible with broad substrate specificity. Indeed, a tyrosine residue from two different Ceg4 tags was found to make intimate contacts with the active site of a neighboring molecule in two different crystal lattices (Figure 2.19). This was taken to be an indication of substrate promiscuity allowed by the Ceg4 cap domain.

Ceg4 likely exerts its effect at a late stage of infection, as evidenced by low transcription levels prior to exponential growth of *L. pneumophila* (Quaile et al., 2018). These studies support the role of Ceg4 as a HAD-like phosphotyrosine phosphatase and provide an example of substrate promiscuity afforded by a novel cap domain.

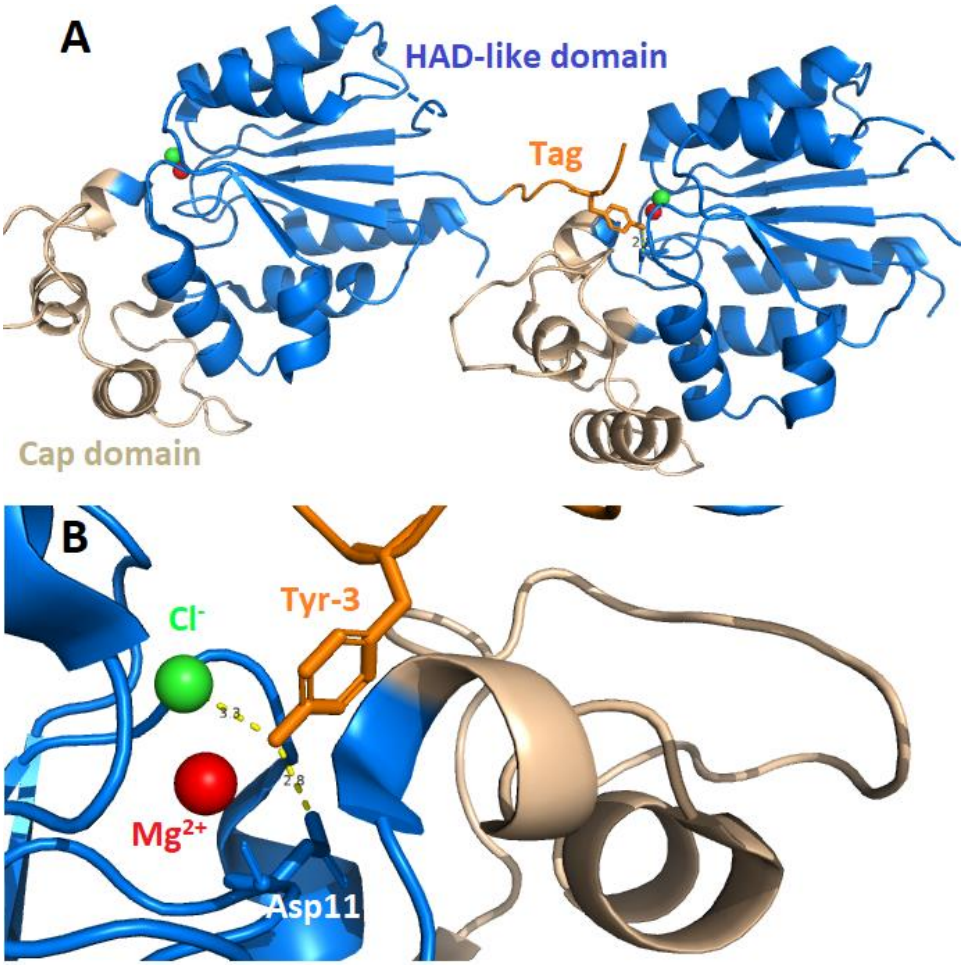

Figure 2. 19 Crystal structure of Ceg4. (A) The expression tag (orange) of one chain extends past the cap domain (grey) and into the catalytic pocket of another. (B) A tyrosine residue in the tag hydrogen bonds with Asp11 and a chloride anion in the HAD-like domain of a neighboring chain. Figure adapted from (Quaile et al., 2018).

Legionella also encodes SH2-domain containing effectors that bind pTyr with much higher affinity than eukaryotic SH2 proteins. Crystal structures have been solved for two of these pTyr super-binders in complex with tyrosine phosphorylated peptides: 1) *L. longbeachae*

LeSH and 2) *L. pneumophila* RavO. Compared to the Src SH2 domain, LeSH contains a 45-residue insertion called the pancreatic polypeptide (PP) that undergoes conformational changes upon peptide binding. This unique clamping mechanism may help to explain the variability in phosphorylated peptides bound by *Legionella* SH2 domains, as eukaryotic SH2 domains respond differently to substrate binding. More importantly, however, was the finding that LeSH lacks a specificity pocket for sidechains near the phosphorylated tyrosine. Although RavO exhibited modest preference for hydrophobic residues in the P + 1 position, it too accommodated a wider variety of phosphorylated peptides than its eukaryotic counterparts. The presence of bacterial SH2-domain containing proteins lacking distinct specificity pockets may suggest that these proteins have evolved independently of tyrosine kinases, which also show preference for specific sequences (Kaneko et al., 2018).

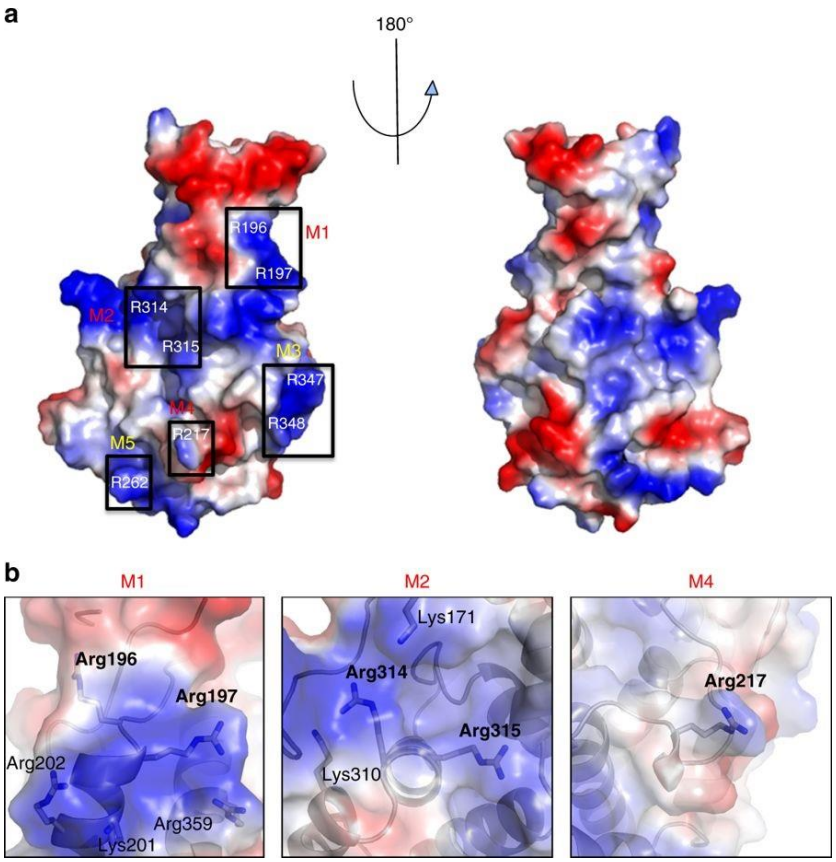

Figure 2. 20 The effector-recruiting surface of DotM. (A) Crystal structure of DotM shown as a surface electrostatic representation. Putative E-block binding residues are labelled and divided into regions (M1 – M5). (B) Essential residues for effector binding are shown as sticks. They belong to M1, M2 and M4 regions. Figure adapted from (Meir et al., 2018).

2.2.2.6. Structural components of the type IVB secretion system

L. pneumophila injects many of its effectors into the host cell using an envelope-spanning multiprotein complex called the intracellular multiplication / defect in organelle transport (Icm/Dot) type IVB secretion system (T4BSS). This functionally diverse nanomachine shares evolutionary roots with the bacterial conjugation apparatus, which can translocate a wide variety of substrates out of the cell, including DNA, toxins and biofilm components. By contrast, the *L. pneumophila* T4BSS is a dedicated effector translocating machine that recruits substrates by two distinct mechanisms: E-block mediated and IcmSW-dependent translocation.

A recent crystal structure of the *L. pneumophila* DotM protein provides insight into the E-block mediated approach of effector uptake. DotL, DotM and DotN form the T4BSS coupling complex, which recruits effectors and funnels them into the secretion channel. While DotL is exclusively an inner membrane protein, both DotM and DotN have cytoplasmic components that may be involved in effector recognition (Meir et al., 2018). The DotM structure reveals several patches of basic residues (M1 – M5) that may bind the glutamate-rich motif (E-block) of secreted effectors (Figure 2.20). Mutagenesis of residues in the M1, M2 and M4 regions completely abolished the interaction of DotM with effectors, as evidenced by isothermal titration calorimetry. These studies pinpoint a set of crystal effector-binding residues within the DotM. Presumably, E-block-containing effectors dock to these basic residues on DotM prior to entering the secretion channel (Meir et al., 2018).

IcmS and IcmW are a set of chaperones that maintain effectors in a conformation that is amenable to translocation. Until recently, it was unclear if the IcmSW heterodimer functioned independently or as a component of the T4BSS. X-ray crystallography has brought clarity to this issue by establishing that IcmSW is an integral part of the T4BSS coupling complex. Indeed, the crystal structure of a complex containing IcmSW and DotL(656-783) showed the latter to adopt an extended conformation spanning IcmSW. This IcmSW-DotL(656-783) structure contains a concave surface lined with hydrophobic residues, which was found in another crystal structure to be the binding site for an additional component of the core complex called LvgA. By docking crystal structures of DotL(590-569)-DotN and DotL(656-783)-IcmSW-LvgA onto a small-angle X-ray scattering (SAX) envelope, individual components of the ring-like T4BSS coupling complex were determined (Figure 2.21). These components were

used to generate a hexameric assembly, which agrees with the likely oligomeric state of DotL (Kwak et al., 2017). These data did not point to a clear role for IcmSW in facilitating the translocation of effectors, however. Additional studies using photo-crosslinking assays showed IcmSW to interact with effectors via the DotL binding interface. In accordance with these findings, neither IcmSW-DotL, IcmSW-LvgA nor IcmSW-DotL-LvgA were able to bind effectors. These results point to a dual role for IcmSW in binding the T4BSS and effector proteins (Xu et al., 2017), although further studies are needed to reveal how these exclusive binding events are coordinated.

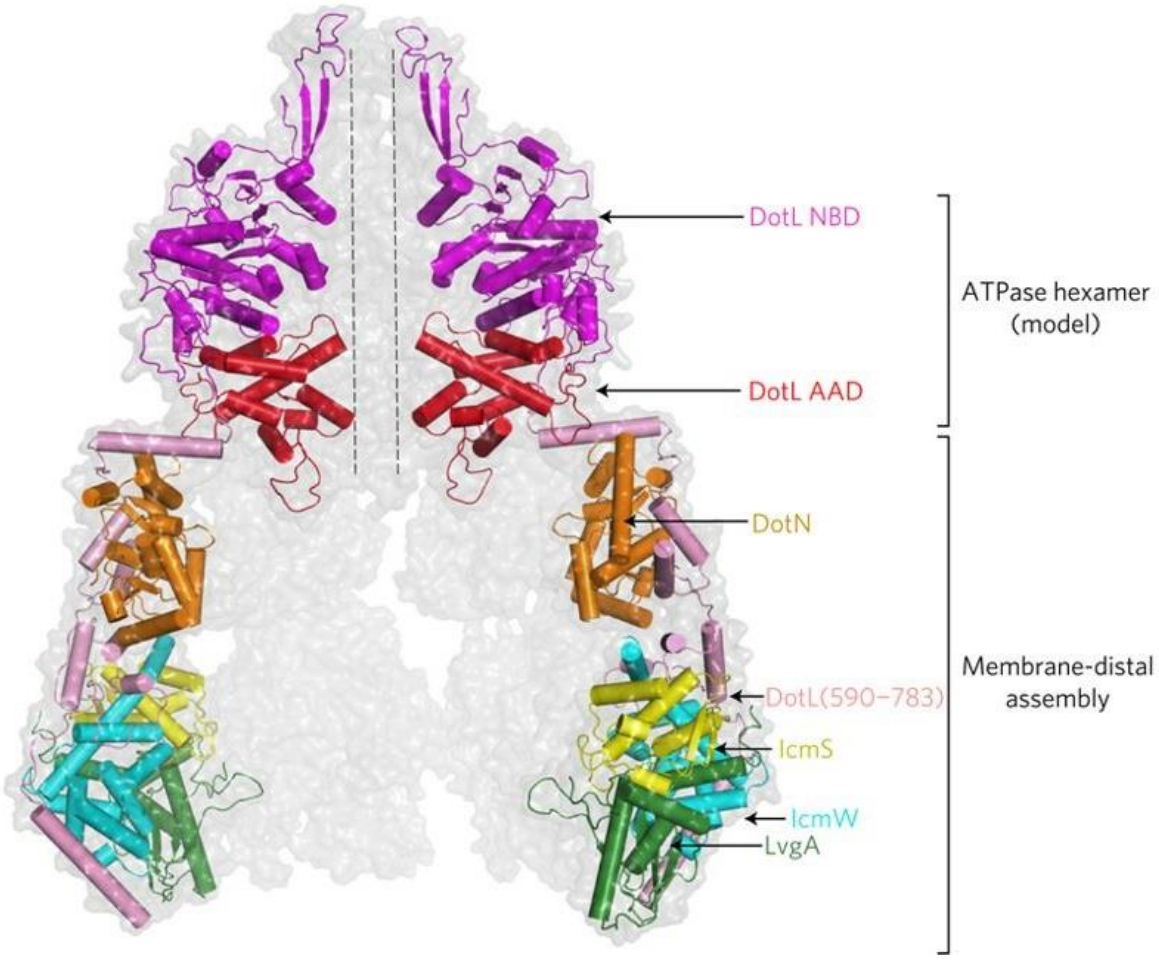

Figure 2. 21 Model of the hexameric type 4 coupling protein (T4CP) complex showing two opposing monomers. DotL (pink) adopts an extended conformation that spans DotN (orange) and IcmSW (yellow and cyan). Figure adapted from (Kwak et al., 2017).

Another component of the T4BSS is an AAA+ ATPase called DotB, which is a major energy source for the system. This protein is thought to facilitate assembly of the T4BSS and promote secretion of substrates. ATPases are generally understood to cycle through three conformations during catalysis: apo, ATP-bound and ADP-bound. Surprisingly, a crystal structure of the DotB hexamer from *L. pneumophila* ($DotB_L$) showed subunits to take on two unique conformations that do not depend on the state of nucleotide binding. Indeed, $DotB_L$ was crystallized in the presence of AMP-PNP and a phosphate was seen in the active site of both conformers. $DotB_L$ is comprised of an N-terminal PAS-like domain and a C-terminal domain containing Walker A and B motifs. By aligning the N-terminal domains of $DotB_L$ proteins comprising the hexamer, two distinct orientations of the C-terminal domain were observed.

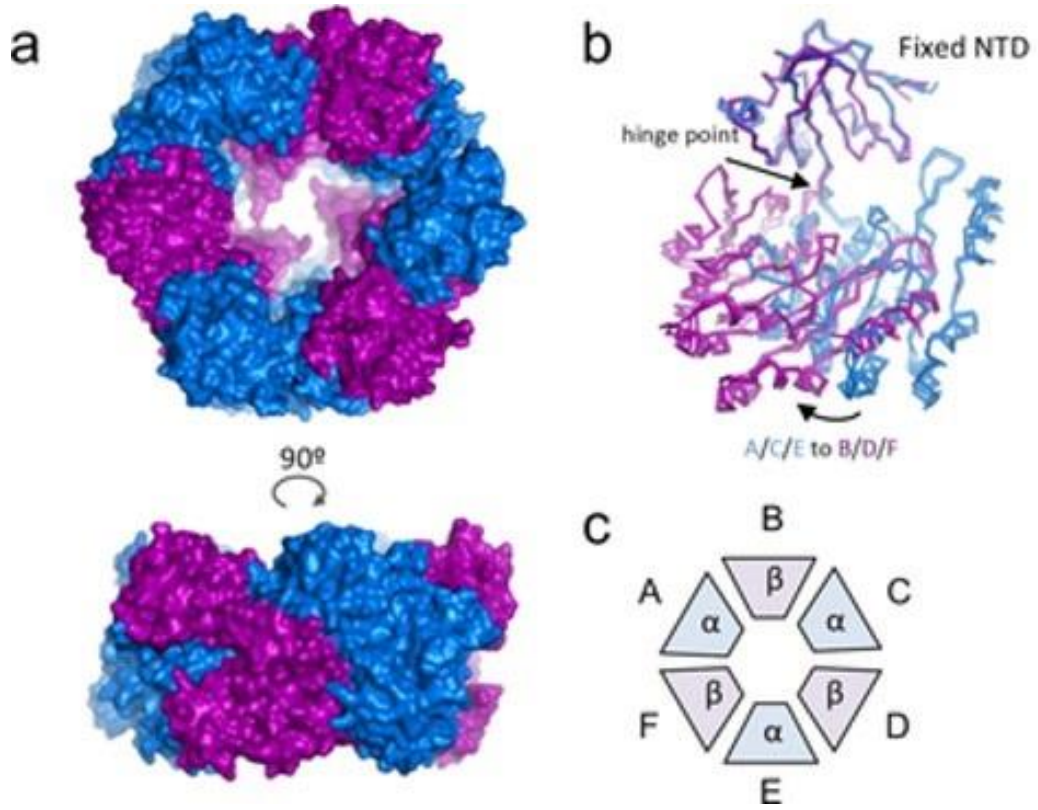

Figure 2. 22 Crystal structure of the DotB hexamer. (A) Top and side views of the DotB hexamer with α and β conformers shown in blue and purple, respectively. (B) Both DotB conformers aligned by their N-terminal domain (NTD) to show a hinge point giving rise to unique C-terminal domain (CTD) orientations. (C) Organization of conformers in DotB hexamer, as viewed from above. Figure adapted from (Prevost and Waksman, 2018).

These unique conformations of $DotB_L$ were called α and β, with the β conformer having its C-terminus rotated 46° toward the center of the hexamer (Figure 2.22). Since α and β conformers occurred alternately in the $DotB_L$ hexamer (αβαβαβ), the rotation of the C-terminus in β produced unique contact interfaces within and between subunits. In contrast to $DotB_L$, the hexamer of *Yersinia pseudotuberculosis* ($DotB_Y$) adopts an ααβααβ organization. Although $DotB_Y$ was crystallized in the absence of ATP, structural analysis showed the β conformer to have an active site that accommodates phosphate binding, while the α conformer does not. This was evidenced by an unwinding of the first helix in the Walker A motif in only the β conformer of $DotB_Y$. Taken together, these structural findings reveal a new level of conformational complexity for DotB proteins (Prevost and Waksman, 2018). Effectors involved in nucleotide metabolism

An essential property of many eukaryotic proteins is their ability to bind nucleotides. For example, kinases rely on an ATP-binding site to facilitate the transfer of phosphate moieties onto a target protein. Another important class of nucleotide-binding molecules are G-proteins, which regulate vesicular trafficking and bind GDP or GTP to transition from inactive to active states, respectively. Furthermore, nucleotides such as cyclic-AMP (cAMP) and cGMP act as second messengers in a variety of signaling pathways. Since many *L. pneumophila* effectors have been acquired from eukaryotes by horizontal gene transfer, it is not altogether surprising that several contain nucleotide binding sites, considering the broad role of nucleotides in host cell processes.

While many nucleotide binding effectors have a clearly defined function, as is the case for kinases, phosphatases, glycohydrolases and SidE-family proteins, others remain functionally uncharacterized. One such protein is the *L. pneumophila* effector, lpg1496, which comprises three domains with unique nucleotide-binding capabilities. A crystal structure of the lpg1496 C-terminus reveals a domain homologous to the cAMP-specific phosphodiesterase (PDE) from *Leishmania*, called LmjPDEB1. Although malachite green assays against an array of nucleotides showed no PDE activity for the C-terminal domain of lpg1496, this construct co-crystallized with ADP. The authors suggested ADP-ribosyltransferase activity for this domain, based on the unique orientation of ADP in the active site relative to other known PDEs. Limited

proteolysis on lpg1496 revealed two N-terminal domains with homology to histidine kinase-like ATP binding regions and S-adenosylmethionine-dependent methyltransferase proteins (KLAMPs). Crystal structures were solved for both N-terminal KLAMP domains of lpg1496, revealing two antiparallel α-helices, flanked on either side by solvent exposed β-sheets. Intriguingly, these KLAMP domains have different nucleotide preferences. Specifically, the N-terminal KLAMP domain (KLAMP1) binds 3'-5'-cAMP, whereas the middle KLAMP domain (KLAMP2) binds ADP. These binding properties likely arise from conformational differences at the nucleotide binding β4-β5 loop, which harbors a *cis*-proline in KLAMP2 and a *trans*-proline in KLAMP1 (Figure 2.23) (Wong et al., 2015). The function of these domains and why they occur together in the same protein are questions that remain unanswered. This study reveals a novel and versatile nucleotide binding fold in bacterial effectors that may have unique functional properties.

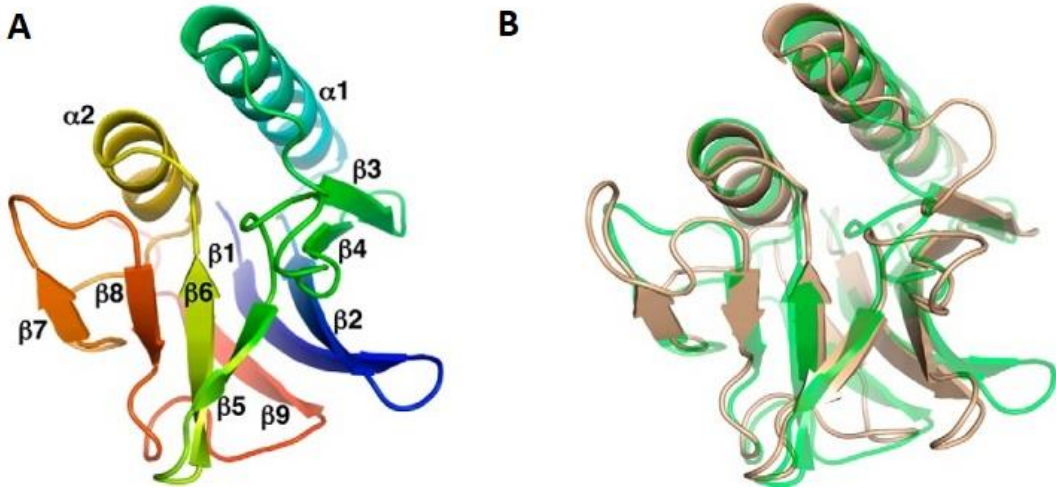

Figure 2. 23 Crystal structure of KLAMP domains in lpg1496. (A) Model of the KLAMP1 domain with secondary structure elements labeled. (B) Overlay of KLAMP1 (faint green) and KLAMP2 (wheat) domains, showing similar overall architecture with differences in the loop regions. Note the conformational differences between β4-β5 loops, which account for the unique ligand binding properties of these domains. Figure adapted from (Wong et al., 2015).

Another nucleotide binding protein found in *L. pneumophila* is the hypoxanthine-guanine phosphoribosyltransferase (*Lp*HGPRT), derived from eukaryotic proteins with the same function. HGPRTs play a role in the nucleotide salvage pathway, which generates inosine or

guanosine monophosphate to be incorporated into DNA or RNA. Supplementing the *L. pneumophila* growth medium with guanine has been shown to increase the replication rate of this pathogen. *Lp*HGPRT may therefore facilitate the nucleotide synthesis that accompanies bacterial growth under these conditions. Crystal structures of apo-*Lp*HGPRT and GMP-*Lp*HGPRT reveal an overall α/β fold like that of 6-oxopurine PRTs. Four loops (I – IV) surround the ligand binding site and are organized by the dimerization of *Lp*HGPRT. Unlike the human HGPRT enzymes, loop II of *Lp*HGPRT is conformationally similar in apo- and ligand-bound structures. This similarity arises from an arginine residue on loop II that hydrogen bonds to the GMP phosphate while stabilizing loop II by interacting with a threonine residue on a nearby α-helix (Figure 2.24). In human HGPRT proteins, this position contains a serine residue that is unable to mediate these stabilizing interactions, explaining the conformational flexibility observed in loop II of human HGPRTs. The fact that loop II is clearly defined in the electron density of both apo- and GMP-bound *Lp*HGPRT suggests that this loop is stably anchored into position to cover bound nucleotide (Zhang et al., 2016). By limiting the motions required to achieve an active conformation, this single mutation increases the efficiency of *Lp*HGPRT.

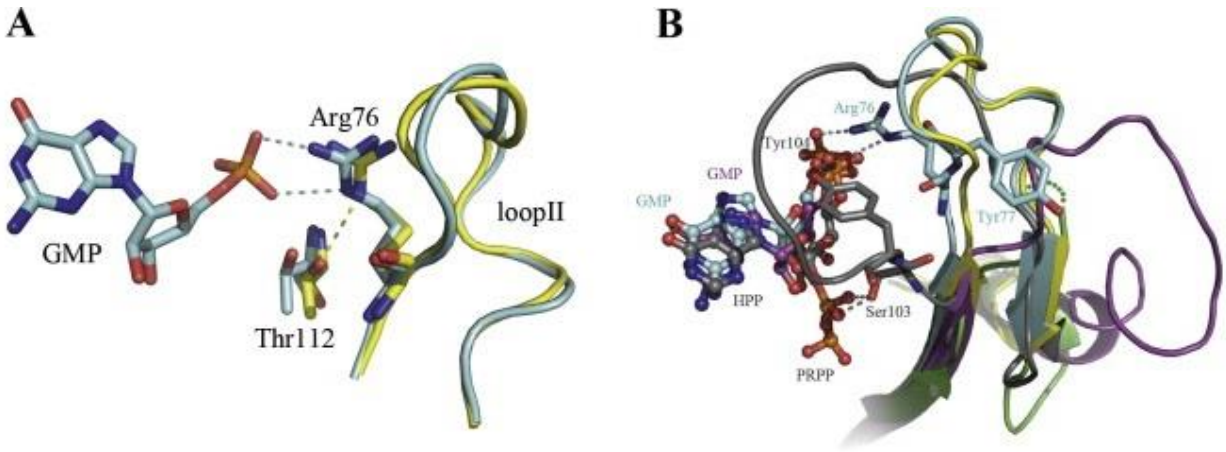

Figure 2. 24 Crystal structure of LpHGPRT. (A) Conformation of loop II in apo- (yellow) and ligand-bound (cyan) LpHGPRT. Note the stabilizing interactions contributed by Arg76 in both models. (B) Overlay of apo-human HGPRT (PDB id 1Z7G) (green), human-HGPRT-GMP (PDB id 1HMNP) (purple) and K68A HGPRTase HPP PRPP Mg complex (PDB id 1D6N) (grey). LpHGPRT is coloured as in (A). Figure adapted from (Zhang et al., 2016).

3. Rationale and Objectives

Legionella pneumophila causes Legionnaire's disease predominantly in elderly or immunocompromised individuals. This pathogen has an unusually large number of effector proteins. Indeed, the effector repertoire of *L. pneumophila* is nearly five times that of *Escherichia coli* O157:H7 and *Salmonella enterica.* It is well adapted to a large spectrum of protozoa (Ensminger, 2016) and has accumulated a wide range of mechanisms to overtake these varied hosts. Studying such infective tactics can expose how bacterial pathogens evade the eukaryotic immune response. *L. pneumophila* can also be studied in mouse models to observe the way higher eukaryotes react to infection. Another advantage of *L. pneumophila* is its intermediate virulence level toward humans. Since milder pathogens do not respond as quickly to the host immune system, researchers use these organisms to observe aspects of the host/pathogen interplay that would not otherwise be detectable.

The virulence of intracellular pathogens is characterized by their effector protein repertoire, which allows them to establish a replicative niche inside host cells and evade xenophagy (Kwon and Song, 2018). Most effectors are substrates of the Icm/Dot type IV secretion system (T4SS) and are injected into the host cell during infection. Each *Legionella* species encodes a unique complement of effectors with only eight core T4SS substrates being present in all genomes. This variability in effectors suggests that *Legionella* species have acquired different virulence strategies during their evolution. The predicted number of effectors has recently been expanded to include additional proteins with eukaryotic-like motifs and their orthologs. Together, nearly 18000 unique effectors are encoded by the *Legionella* genus (Gomez-Valero et al., 2019). The study of *Legionella* virulence is therefore a formidable undertaking that will reveal many ways in which eukaryotic processes are coopted by the pathogen.

Functional redundancy is an obstacle in characterizing effector proteins; deletion of a gene encoding a redundant effector is unlikely to produce a phenotype (Ensminger, 2016). Structural data on such a protein can provide an alternative avenue toward functional analysis. For example, protein structures can be used to predict putative catalytic residues, allosteric regulators and regions of functional significance on the protein of interest.

The work in this dissertation began with a broad screen of uncharacterized *L. pneumophila* effectors for their suitability in structural studies. Several proteins were promising, due to their

level of expression, solubility and homogeneity. Of these, LpnE, MavL and MavE produced well-diffracting crystals. I was able to solve their structures and perform some functional characterization on each of these proteins. LpnE had been partially functionally characterized at the onset of the project. Conversely, there was little published data available for MavL and MavE. These proteins are actively being investigated by our collaborators and their three-dimensional structure will be of help in directing their efforts toward full functional characterization.

LpnE is an effector implicated in host cell entry that is exported independently of the T4- or T2SS (Newton et al., 2006, 2007a). Once in the macrophage cytosol, LpnE is anchored to the *Legionella*-containing vacuole (LCV) and binds the N-terminal 236 residues of oculocerebrorenal syndrome of Lowe (OCRL) protein. Recruitment of OCRL to the LCV has been found to restrict the intracellular replication of *L. pneumophila* by an unknown mechanism (Weber et al., 2009). The objective of this project is to obtain molecular insights on the interaction of LpnE with OCRL using structural biology and protein-protein interaction studies. In so doing, we hope to expand our understanding of Sel1-like repeat (SLR) proteins in pathogenicity.

MavL is a substrate of the T2- and T4SS, with homologues in all *L. pneumophila* species but none in *L. longbeachae* (DebRoy et al., 2006; Gomez-Valero et al., 2011). There is presently little data on MavL in literature, although our collaborators Elizabeth Hartland et al. have found MavL to bind the E2 ubiquitin conjugating enzyme, Ube2q1/Nice5, by yeast two-hybrid and co-immunoprecipitation (unpublished data). The potential role of MavL in host ubiquitination remains to be elucidated. The objective of this project is to obtain a crystal structure of MavL as a step toward understanding its role in *L. pneumophila* pathogenesis.

MavE is another effector of unknown function translocated by the T4SS (Huang et al., 2011). Our collaborator, Dr. Yousef Abu-Kwaik, has discovered that *L. pneumophila* lacking MavE form LCVs that are targeted to lysosomes for destruction (unpublished data). Thus, MavE appears to play a significant and non-redundant role in LCV trafficking and/or maturation. The objective of this project is to characterize MavE by X-ray crystallography. In so doing, we hope to gain molecular insights into its function in vacuolar trafficking.

4. Utilized Methods

This chapter describes the general methods employed in all projects outlined in this dissertation. In sections 4.1 – 4.6, the protocols for protein cloning, expression, purification and analysis are given. Section 4.7 is a description of the methodology and theoretical background of X-ray crystallography. Finally, a list of primers and chemicals used is provided in section 4.8

4.1. Cloning

Protein constructs were designed using the ligation independent cloning (LIC). Briefly, this approach relies on the generation of complementary overhangs on the insert and vector, such that annealing occurs spontaneously and without a ligase enzyme. These overhangs are produced using the 3' → 5' exonuclease activity of T4 DNA polymerase in the presence of either dGTP (for the vector) or dCTP (for the insert). The T4 exonuclease is unable to hydrolyze deoxy-nucleotides and therefore halts when it encounters dGTP or dCTP in the vector or insert, respectively. This leads to the production of 15-base pair complementary overhangs, which can anneal in the absence of ligase (Eschenfeldt et al., 2009). By this approach, the DNA sequences were cloned into the pMCSG7 expression vector (Stols et al., 2002) incorporating a TEV-cleavable, N-terminal His_6-tag.

4.2. Protein Expression

The following is a general protocol used in the large-scale expression of recombinant proteins described in this dissertation. Plasmids were transformed into chemically competent BL21(DE3) cells and plated on LB agar containing ampicillin (100 µg/mL). A single transformant was inoculated into 20 mL of LB supplemented with ampicillin (100 µg/mL) and glucose (0.4%), and grown overnight at 37 °C. This overnight suspension was subcultured into 1 L of terrific broth (TB) supplemented with ampicillin (100 µg/mL) and grown at 37 °C. Once the cell culture reached an optical density (A_{600}) of ~1.0, the temperature was reduced to 18

°C, 1 mM of isopropyl β-D thiogalactopyranoside (IPTG) was added to the culture to induce protein expression and the cells were incubated for ~16 more hours. Cells were pelleted at 6900 g for 15 min in a Beckman JLA 8.1000 rotor and stored at -80 °C until further processed. Approximately 10 g of pellet was obtained from 1 L of culture.

4.3. Protein Purification

Cells were resuspended in 30 mL of a lysis buffer (50 mM Tris, pH 8.0, 10% (v/v) glycerol, 0.1% (v/v) Triton X-100) and lysed twice at 35 kPsi in a cell disruptor (Constant Cell Disruption Systems, Kennesaw, Georgia). The lysate was spun at 21 000 g for 30 min in a Beckman JA25.50 rotor. Supernatant was added to 5 mL of Qiagen Ni-NTA beads preequilibrated with three column volumes of a buffer containing 20 mM Tris, pH 8.0, and 50 mM NaCl, and the beads were washed with 50 mL of this buffer. Protein was eluted with the same buffer supplemented with 100 mM imidazole. Purified protein was concentrated to 16 mg/mL in a Millipore centrifugal filter spun at 4000 g. In the final step of purification, the protein was loaded onto a Biorad or GE gel filtration column to isolate the monomeric or oligomeric species of interest.

4.4. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a method for separating proteins based on their molecular weight (Shapiro et al., 1967). The speed at which a protein moves within an electric current depends on its charge-to-mass ratio. If the charge of all proteins examined is brought to a constant, however, the speed of protein migration can be related directly to its mass. This equalization of charge between proteins can be achieved using sodium dodecyl sulfate (SDS), a strong anionic detergent that binds and linearizes protein to form a denatured protein-detergent-complex (PDC). Importantly, the PDC contains approximately half the number of SDS molecules as there are amino acids in the protein (Smith, 1994). Since each molecule of SDS is negatively charged, the overall charge of the PDC is overwhelmingly negative. Thus, solubilizing a protein in SDS minimizes differences

in charge between proteins, rendering mass the only remaining parameter to distinguish them (Shapiro et al., 1967; Smith, 1994). The PDCs then migrate through a sieve-like polyacrylamide matrix under an electric field (Smith, 1994).

SDS-PAGE separates proteins by the method of isotachophoresis. In this approach, stacking- and resolving-polyacrylamide gels of differing pH are used. The protein sample initially passes through the stacking gel, where it is concentrated between leading (chloride) and terminating (glycinate) ions present in the reservoir buffer. In the resolving gel, the mobility of glycinate overtakes that of PDCs due to a pH change. Proteins can then be separated from one another in the resolving gel based on size, because they are no longer concentrated between chloride and glycinate ions (Smith, 1994).

4.5. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a method used to measure the heat taken up (endothermic) or expelled (exothermic) during a chemical reaction. These values reflect the extent of a reaction (in moles) and its change in enthalpy (ΔH). ITC can also be used to measure the rate of a reaction, which is proportional to the rate at which heat is exchanged with the surroundings. Thus, ITC is a valuable tool for determining the amount of a reaction taking place and the rate at which a reaction occurs (Freyer and Lewis, 2008).

An isothermal titration calorimeter measures the enthalpy of a reaction by the method of power compensation. In this approach, the calorimeter measurement cell is maintained at constant temperature using a control heater which responds to changes in heat by varying its power output. A neighboring cell in the instrument (containing only reaction buffer or water) is supplied with constant power by the control heater to serve as a reference for temperature changes in the measurement cell. The raw output in the ITC experiment is the power (µJ/s) required to keep the measurement cell at constant temperature (Freyer and Lewis, 2008).

ITC experiments are typically conducted as a series of injections, where analyte (ligand) is titrated into the measurement cell containing the protein of interest. Selection of appropriate concentrations for ligand and protein is a critical factor in the experiment. Excessive analyte could lead to a rapid saturation of binding sites on the protein, while too little analyte or protein

may produce undetectable heat changes. In practice, the ligand is often kept at a greater concentration than the protein to ensure that saturation of binding sites occurs during the experiment. This is because the volume of protein in the measurement cell (350 µL) far exceeds that used for individual titrations (5 µL). Therefore, high analyte concentrations are often needed to achieve the appropriate reaction stoichiometry. Saturation of ligand-binding sites is reflected by successive decreases in the measured heat changes. In such a case, the largest heat change in the experiment represents a case where all binding sites in the protein are occupied by ligand. If the protein/ligand interaction is reversible, we expect some unbound protein to be available for ligand binding in the next injection. This is evidenced by a smaller heat change because only a subset of analyte molecules interact with the ligand. Heat changes continue to diminish in subsequent injections until effectively no protein is available to bind analyte.

ITC is an extremely sensitive method for measuring changes in heat. Even the titration of water into water gives a measurable signal. As such, a background exists for all injections measuring the interaction of a ligand with protein. To account for this, an additional experiment should be carried out where buffer is titrated into the same solution of protein as that used in the protein/ligand experiment. Importantly, the injected buffer should be the same as that used to prepare the ligand This is because even slight differences in buffer composition can affect the measured signal.

In summary, ITC quantifies the heat given off or absorbed during a chemical reaction. This method is particularly useful for measuring interactions between a protein and ligand. By monitoring the change in heat during successive injections of analyte into protein, information on binding stoichiometry and even a binding constant can be obtained. However, a profile of buffer into protein should be obtained and subtracted from the raw protein/ligand data to correct for background heat changes that do not reflect the biological phenomenon under study.

4.6. Dynamic light scattering

Dynamic light scattering (DLS) is a method for estimating the size distribution (polydispersity) of molecules in solution. In protein biochemistry, DLS is used to determine the homogeneity and relative size of a molecule. Aggregated or precipitated protein is also

readily detected by this method. Taken together, these properties make DLS an excellent quality assessment tool for purified protein prior to performing crystallization trials or biological assays (Maguire et al., 2018).

The physical principle underlying DLS is that the hydrodynamic diameter of a molecule in solution can be related to its diffusion coefficient. Since molecules in solution undergo Brownian motion, the intensity of scattered light will fluctuate over time. The decay in scattered light intensity is used to generate an autocorrelation curve, which is proportional to the diffusion coefficient of the molecule under study (provided the sample is free of contaminants). Once obtained, the sphere-equivalent hydrodynamic diameter of the molecule can be calculated using equation 4.6:

$$D_t = Tk_B/3\pi\eta d \qquad 4.6$$

where D_t is the diffusion coefficient, T is the temperature in Kelvin, k_B is Boltzmann's constant, η is the solution viscosity and d is the hydrodynamic diameter of the molecule. It is important to note that all molecular weights calculated from this value of d assume a spherical molecule in solution. Thus, size estimates obtained from DLS are only approximations of the true molecular weight (Maguire et al., 2018).

4.7. X-ray crystallography in macromolecular structure determination

4.7.1. A brief overview of crystallography

In 1895, Röntgen found that the permeability of an object to X-rays could be related to its density (Röntgen, 1896). Laue expanded on this finding by suggesting that radiation diffracts when passed through a crystal, provided the wavelength of this radiation is similar in magnitude to the distance between lattice planes. This hypothesis was verified by Friedrich and Knipping, who showed that a diffraction pattern is indeed produced when copper sulfate crystals are irradiated with X-rays (Friedrich and Knipping, 1913). Bragg then described the diffraction process as reflection of partial rays from a surface (Bragg, 1912). A critical property

of this surface is periodicity, since even slight irregularities would abolish the reflected wave magnitude. Crystals are characterized by symmetry and consistent, repeating units. Thus, when a crystal is irradiated with X-rays, constructively interfering waves are amplified such that diffraction maxima are visible on a detector (Rupp, 2010).

The description of diffraction as reflection of waves from a surface allowed Bragg to develop an important mathematical relationship describing the condition for constructive interference called Bragg's law (equation 4.1), which states that the path difference between two waves must equal an integer number of wavelengths if constructive interference is to occur (Bragg, 1913).

$$2dsin\theta = n\lambda \qquad 4.1$$

Relating interplanar spacing to incident wave angle allowed a deeper connection between the diffraction pattern and crystal content to be made. First, the sum of all constructively interfering waves reflected from a common set of planes was termed a structure factor. As such, each diffraction maxima could be assigned a structure factor corresponding to a specific set of planes (*hkl*) in the crystal (equation 4.2). Importantly, the structure factor is a Fourier integral of electron density within the crystal. Since the Fourier transform is invertible, another Fourier integral also exists wherein electron density is described as a summation of structure factors representing the diffraction pattern (equation 4.3) (Rupp, 2010). Thus, a mathematical description of electron density within the crystal can be uncovered by performing an inverse Fourier transform on the complex structure factors. This description of electron density is often rendered as a contour map, which serves as a guide for initial model building.

$$F_{hkl} = \sum_{j=1}^{n} f_j \, e^{2\pi i(hx_j + ky_j + lz_j)} = |F_{hkl}|e^{i\phi} \qquad 4.2$$

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} F_{hkl} e^{-2\pi i(hx_j + ky_j + lz_j)} \quad 4.3$$

We measure the intensity I of the scattered radiation, which is proportional to $|F_{hkl}|^2$ but do not directly measure the phase angle ϕ (eq. 4.2). Therefore, the phase of the structure factor is not recorded during data collection for each measured reflection. The lack of phase information, which is an essential component of a structure factor, is called the "phase problem." Several experimental and computational methods have been developed to determine the missing phases. These approaches are described briefly in subsequent sections.

In summary, crystals are periodic structures that allow the visualization of diffraction maxima, which occur only where Bragg's law for constructive interference is met. Constructively interfering waves (and their experimentally determined phases) are combined into a Fourier integral called a structure factor, which represents diffraction from a set of planes within the crystal. The inverse Fourier transform of all structure factors gives a mathematical description of electron density distribution that can be viewed as a contour map outlining the unknown molecule (Rupp, 2010).

4.7.2. Growing protein crystals

Crystallizing a protein requires that a super-saturated state is achieved without precipitating, denaturing or otherwise altering the molecule of interest (Moreno, 2017). This is done by adding mild precipitating agents and changing other parameters such as temperature or pH. In addition, ligands or ions that affect the protein conformation may be required to ensure homogeneity for crystallization. Under ideal conditions, these mild precipitating agents will concentrate the protein in such a way that crystallization occurs, while other forms of coalescence such as precipitation or phase separation are avoided (Chernov and Komatsu, 1995; Malkin et al., 1996).

The process of crystallization is divided into nucleation and growth. In nucleation, molecules randomly distributed in solution assemble into an ordered array. This initial stage of order is followed by the growth phase, wherein additional molecules of the same kind organize themselves onto the nuclei scaffold. Excessive nucleation can arise when supersaturation levels of the protein are too high. Such conditions may lead to showers of microcrystals or precipitation (Chernov and Komatsu, 1995; Malkin et al., 1996). Alternatively, no nucleation events will occur if the protein concentration is too low. Arriving at the optimal level of supersaturation for nucleation is therefore critical for ensuring proper crystal formation and growth conditions.

Supersaturation is a state where protein remains in solution despite being present at a concentration beyond the solubility limit. To reestablish equilibrium, excess protein may form aggregates, precipitates or crystals. A phase diagram can be used to describe the critical points of saturation with respect to protein and salt concentration (Figure 4.1). On this diagram, a clear boundary distinguishes the stable and metastable states. This is because the solubility limit for a protein under specific conditions can be readily defined. The metastable state represents the low end of supersaturation, wherein crystal growth can occur, but nucleation is unlikely. Conversely, both nucleation and growth can occur in the labile state (Moreno, 2017). Ideally, we would hope to have our protein enter the labile level of supersaturation initially, followed by a lengthy period in the metastable zone. This is possible, as the formation of nuclei from existing protein in solution may indeed reduce the saturation level from labile to metastable. Still, the range of conditions under which such an event can occur are limited. Comprehensive screening of crystallization conditions and optimization is therefore needed to find the right condition (Garman, 2014).

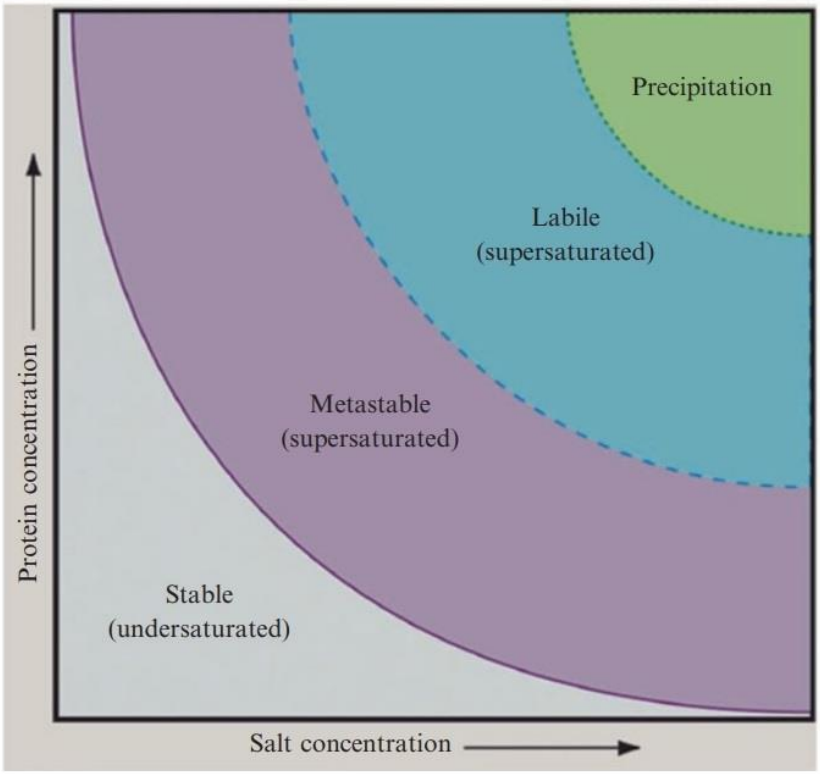

Figure 4. 1 Phase Diagram for Protein Crystallization. Under conditions of low salt and protein concentration, the solution remains undersaturated. The solubility of protein in solution can be gradually decreased by raising the salt concentration. In the metastable state of supersaturation, protein crystals may grow but are unable to nucleate. Both nucleation and crystal growth may occur in the labile state of supersaturation. At very high salt and protein concentrations, the protein is likely to crash out of solution in a precipitated form. This Figure is adapted from (Moreno, 2017).

The nature of the protein itself is also a critical factor in the crystallization process. For example, disordered or flexible regions on a protein may obstruct potential crystal contacts. Recombinant DNA technology has afforded researchers the luxury of stripping those unwanted regions to produce a new protein construct that is amenable to crystallization. Co-crystallization with antigen-binding fragments (Fabs) may also facilitate crystal growth by generating intermolecular contacts that would not otherwise be present (Bailey et al., 2018). This strategy is particularly beneficial in the case of membrane proteins, which are often solubilized in a detergent micelle that increases the distance between regions of contact.

Other methods of altering the protein to increase the likelihood of crystallization include methylation, surface-entropy reduction and limited proteolysis. Methylation of surface lysine

residues reduces protein solubility and makes it easier to achieve a supersaturated state. Surface entropy reduction depends on computer-assisted prediction of surface residues that may interfere with crystal formation (Sledz et al., 2010; Walter et al., 2006). Residues that increase surface entropy are often those with long and flexible sidechains, such as lysine or arginine. Identification of the putative interfering residues is followed by site-directed mutagenesis, expression, purification and crystallization trials (Derewenda and Godzik, 2017). Finally, limited proteolysis can be an efficient way to arrive at an appropriate construct for crystallization. For example, any stable product arising from digestion of a protein is relatively inaccessible to proteolysis. Since proteases initiate peptide hydrolysis on linear sequences, the stable core is stripped of any such sequences vulnerable to attack. By extension, this core fragment is less likely to harbor flexible regions and may therefore be a better candidate for crystallization. One caveat to this approach is the possibility that regions cleaved by a protease are of functional significance to the protein of interest.

4.7.3. Cryo-crystallography and data collection

Cryo-crystallography is a branch of science that applies crystallographic methods at low temperature (Macchi, 2012). This method was first adapted to the crystals of macromolecules by Häkon Hope (Hope, 1988). The study of crystals at low temperature is often carried out using X-ray diffraction, although other approaches may also be used. Moreover, this science is not limited to structure determination. For example, the study of phase transitions can enhance our understanding of the phase diagram for a crystalline material (Macchi, 2012). This information may be useful for chemists or physicists studying molecular trends under wide temperature ranges. For many interested in protein structure, however, cryo-crystallography has been reduced to a technique used for preserving crystals prior to data collection. This technique has nontrivial advantages for the diffraction experiment, such as decreasing radiation damage, improving resolution and reducing thermal motion (Larsen, 1995).

In macromolecular crystallography, the cryo-technique involves flash freezing a crystal in liquid nitrogen and maintaining it at ~100K for data collection. To ensure that a crystal survives the cooling process, it must first be submerged in a cryoprotectant solution. Common cryoprotectants include glycerol, 2-methyl-2,4-pentanediol and ethylene glycol. These

molecules interact with water and prevent the formation of ice crystals at low temperature. Although water does enter a solid state under these conditions, it is said to be vitrified rather than frozen (Rajan and Matsumura, 2018).

For data collection, the crystal is rotated within the X-ray beam and intensities are recorded on the detector. These intensities are integrated over a small rotation angle to produce images (frames) containing the diffraction data. Critical parameters to monitor during data collection are completeness and redundancy. Completeness gives the ratio of observed to expected reflections, while redundancy tells us how many times a reflection was measured. The number of expected reflections is known for each type of crystal symmetry (Laue group) which, in turn, can be determined using only a few diffraction frames (Wlodawer et al., 2013). Knowledge of crystal symmetry arms us with the ability to measure completeness and redundancy during data collection. Moreover, it allows us to determine the minimum overall rotation needed to collect a full dataset. This can be a critical step in the process, as crystal decay is a common problem during data collection and the lifespan of a crystal is never known *a priori* (Krojer et al., 2013).

4.7.4. Data processing

Processing of X-ray data collected by the rotation method is divided into indexing, integration and scaling. Typically, one of four software packages is used to carry out these processes: Mosflm, HKL2000, the XDS suite and DIALS (Powell, 2017). The automation of data processing was facilitated, in part, by standardizing experimental parameters for both the crystal and instrument.

Indexing begins with the assignment of *hkl* indices to each measured reflection. This is done by projecting the two-dimensional coordinates of each reflection onto scattering vectors of the three-dimensional Ewald sphere. These scattering vectors can be represented as points in the reciprocal lattice (Ewald, 1969). Each reflection is therefore related to a reciprocal lattice point which, in turn, is defined by a set of planes within the crystal (*hkl*). Scattering vectors are then used to search for a unit cell described by the diffraction pattern. Symmetry of the unit cell is defined using 1 of 14 Bravais lattices. Correct assignment of Bravais lattice symmetry allows the location of reflections to be predicted. If there is poor agreement between predicted

and experimental reflections, it is unlikely that the Bravais lattice was assigned correctly (Powell, 2017).

Integration is the process of measuring diffracted intensities by adding together all values for pixels forming a spot. The difficulty in this task is defining a boundary for each reflection. This is a crucial step in processing, as it allows the signal to noise ratio for each reflection to be maximized. Since intensity profiles for each spot are smooth curves, the assignment of spot boundaries is somewhat arbitrary. Several automated methods have been developed to overcome this problem. For example, background pixels can be assigned by systematically removing intense spots until the statistical distribution of remaining spots agrees with that expected for a random sample. Another method used in the Mosflm package is called 'profile tolerance' and relies on the assignment of spot intensities (rather than background). These methods allow summed intensities for each reflection to be obtained (Powell, 2017). As will be seen in the following section, reflection intensities are proportional to structure factor amplitudes, which are ultimately used in the calculation of electron density maps.

Summed intensities obtained in the integration step must be brought to a common scale that accounts for changes occurring during data collection. Radiation intensity, the crystal and even the detector response are some examples of variables in the experiment. Scaling can be thought of as an effort to correct for such variation and produce a set of intensities that are consistent between frames (Powell, 2017). Finally, redundant or symmetry-related reflections are merged to produce the final dataset.

Once processed, a dataset must be assessed for quality. The parameters used for this purpose are redundancy, Rmerge, completeness, CC1/2 and the signal to noise ratio of each reflection. Redundancy refers to the number of times a reflection is measured (see section 4.1.3). Greater redundancy leads to more accurate reflection intensities which, in turn, provide improved structure factors. Rmerge describes the agreement between intensities of symmetry-related reflections. Since these values are expected to be the same, a good dataset should show only limited deviation between them (i.e. a low Rmerge) (Wlodawer et al., 2008). Completeness is the percentage of expected reflections that were measured for each resolution shell. For a protein crystal, completeness of over 75% in the highest resolution shell is considered acceptable, provided the overall value (for all shells) exceeds 95%. CC1/2 is a form

of the Pearson correlation coefficient that evaluates the reproducibility of measurements used to estimate reflection intensity (Wang et al., 2017). It is useful for detecting weak signals in high resolution shells, because it is insensitive to scale factors distinguishing the two frames being compared. Convention dictates that a CC1/2 of over 60% in the highest resolution shell is a good indication of where to cut the data (Karplus and Diederichs, 2012). The signal to noise ratio (denoted $I/\sigma I$) gives the intensity over background for each reflection. Values of 1-2 are typically observed in the highest resolution shell (Evans and Murshudov, 2013).

4.7.5. Phasing methods

Each spot on the detector arises from constructively interfering waves reflected by electrons within the crystal. Thus, every reflection has a corresponding resultant wave that produces it. Crystallographers describe this resultant wave using structure factors. Importantly, structure factors have both amplitude and phase. The amplitude of each structure factor is proportional to the square root of the measured intensity for the associated reflection (Rupp, 2010). Structure factor amplitudes are therefore determined during data collection, whereas phases remain unknown and must be determined indirectly. The process of determining phases for each measured reflection is called the 'phase problem.' Obtaining phases is essential for structure determination, because the electron density map can only be calculated once complete structure factors (having both amplitude and phase) are obtained (Thorn, 2017). This is because electron density is the inverse Fourier transform of structure factors summed over all values of *hkl*. Hence, structure factors and electron density are reciprocally related. Since electron density represents the molecule, the function describing it is said to exist in 'real space' while structure factors exist in 'reciprocal space' (Rupp, 2010).

4.7.5.1. Molecular replacement

Molecular replacement is a method for obtaining initial phase estimates from a homologous model (McCoy, 2017). This requires orienting the homologous structure in the unit cell of the unknown such that phases from the homologue correlate with the experimentally measured intensities. In so doing, estimates of complete structure factors are obtained and used in the

calculation of an electron density map. The orientation process is the greatest computational challenge of molecular replacement and is divided into rotational and translational searches. Various positions of the homologue in the unit cell are evaluated for their agreement with experimental data using Patterson maps calculated for each (Evans and McCoy, 2008; McCoy, 2017). A Patterson map gives interatomic distances and can be calculated without phases (Patterson, 1934). These interatomic distances can be thought of as a barcode for the orientation of each structure in the unit cell. The experimental and model Patterson map are scored for their level of overlap to arrive at the correct placement of the model into the new unit cell (Evans and McCoy, 2008). Finally, complete structure factors are created using amplitudes from experimental data and the phases of our correctly positioned homologue. A molecular replacement solution is characterized by an interpretable electron density map fitted to the homologous model. This model is refined to agree with the experimental data and gradually improve the phase estimates.

4.7.5.2. Single anomalous dispersion

Single anomalous dispersion (SAD) is an experimental approach for determining phases. It relies on the additional scattering that occurs when an atom is irradiated at a characteristic frequency called the absorption edge. This scattering is considered anomalous and leads to differences in the intensities of symmetry-related reflections (Friedel pairs). These differences are used to calculate the positions of anomalous scatterers through the analysis of a difference Patterson map, from which the complete structure factors (F_A) for those atoms giving rise to anomalous scattering can be calculated. This set of anomalously scattering atoms is called a substructure. The scattering amplitudes of the protein and the protein with anomalous scatterers are related by the equation below (equation 4.4).

$$F_{PA} = F_A + F_P \therefore F_P = F_{PA} - F_A \qquad 4.4$$

The scattering amplitudes for the substructure (F_A) can be calculated and the magnitudes of scattering amplitudes for the protein alone (F_P) and protein with anomalous scatterers (F_{PA})

are known from the diffraction experiment. With this information, two possible solutions for the phase of F_P are arrived at using a mathematical operation that is often represented by a Harker diagram (Figure 4.2). This diagram depicts structure factors as vectors on a coordinate system comprising real (x) and imaginary (y) axes. The phase of each structure factor is the angle its vector makes with the positive real axis, while amplitudes are given by the vector length (Rupp, 2010). This notation gives a convenient illustration of equation 4.4 and its application to the phase problem.

In the context of SAD phasing, Harker diagrams are used to depict the difference between symmetry related reflections, or Friedel pairs, during anomalous scattering. To do this, we start with a vector representing a structure factor without anomalous contribution (F_P+), having phase $+\phi$ and magnitude proportional to the vector length. The Friedel mate of this structure factor (F_P-) is the reflection of F_P+ in the real axis, since Friedel pairs have the same amplitude but opposite phase ($-\varphi$). Anomalous scattering has two components, designated *f'* and *f''*, respectively. Importantly, only *f''* contributes to the difference between Friedel pairs (Figure 4.2) and the X-ray wavelength is therefore selected to maximize the anomalous contribution by *f''*. Differences between the structure factor with anomalous contribution ($F_{PA}+$) and its symmetry mate ($F_{PA}-$) can be easily visualized by reflecting $F_{PA}-$ in the real axis (Figure 4.2). By drawing a circle of radius $F_{PA}+$ centred at the head of the *f''*+ vector and another of radius $F_{PA}-$ centred at the head of the *f''*- vector (after reflection in the real axis), we can see two points of intersection corresponding to potential solutions to the phase problem. This problem, called phase ambiguity, forces us to choose an average value between these possible solutions for each structure factor (blue vector, Figure 4.2). Initial phase estimates to be improved upon by density modification are therefore quite inaccurate. Density modification gradually improves these phases using a set of computational approaches to distinguish molecular envelope from solvent. When density modification is applied to the substructure with correct chirality, we can begin to see a sensible electron density map. By improving our averaged phase estimates for each structure factor, density modification simultaneously solves the problem of handedness for our substructure and phase ambiguity. Ultimately, this allows complete structure factors to be obtained for all reflections representing the protein (Hendrickson, 2014).

Importantly, the SAD method of phasing relies on minor differences in scattering between Friedel pairs. Since SAD phasing uses a single dataset, small changes between symmetry related reflections can be measured accurately. By contrast, heavy atom phasing utilizes two or more datasets. This approach depends on significant differences in scattering that are caused by a heavy atom soaked into the crystal. Thus, although SAD phasing is an adaptation of heavy atom methods, it is made possible by minimizing the error in measured reflections.

Unlike protein models, the substructure given by the set of structure factors, F_H, provides no indication of the correct chirality. Common structural features in proteins, such as right-handed α-helices, are evidence that our model has arisen from the proper substructure. Since the substructure is comprised of a small number of atoms, no such features are present within it. That is, we have no way of knowing whether the right- or left-handed substructure is correct *a priori*. This information is reflected by an interpretable electron density map which, in turn, depends on the correct assignment of phases to each structure factor. Phase ambiguity forces us to choose an average value between the possible solutions to the phase problem for each structure factor. Initial phase estimates to be improved upon by density modification are therefore quite inaccurate. Density modification gradually improves these phases using a set of computational approaches to distinguish molecular envelope from solvent. When density modification is applied to the substructure with correct chirality, we can begin to see a sensible electron density map. By improving our averaged phase estimates for each structure factor, density modification simultaneously solves the problem of handedness for our substructure and phase ambiguity. Ultimately, this allows complete structure factors to be obtained for all reflections representing the protein (Hendrickson, 2014).

Light atoms such as nitrogen, oxygen, carbon and hydrogen have absorption edges outside the range of wavelengths used in crystallography. These atoms are common in biological macromolecules but contribute negligibly to anomalous scattering. By contrast, heavy atoms produce a significant anomalous signal. While some proteins contain metals suitable for anomalous phasing (eg. Fe or Zn), others need heavy atoms (eg. Hg or Pt) to be artificially incorporated in a trial-and-error method called soaking (Pike et al., 2016). In this approach, native crystals are transferred to a solution containing the heavy atom, which binds the crystallized protein without disrupting its conformation. Once this occurs, the soaked crystal

can be used to obtain a new dataset with intensity differences caused by the heavy atom. However, any structural changes in the heavy atom derivative protein will also be reflected in the corresponding dataset, provided they do not break the crystal. It is impossible to distinguish these differences from those caused by the heavy atom alone (Rupp, 2010). Thus, several attempts may be required before an isomorphous dataset can be obtained.

The use of selenium circumvents the issues inherent to soaking, as it is incorporated directly into the protein in the form of a nonstandard amino acid called seleno-methionine. Replacing the sulfur atom for selenium has little structural impact on methionine. Thus, seleno-methionine derivatives are typically isomorphous to their native protein counterparts (Hendrickson et al., 1990). A caveat to this method is that the recombinant protein should have enough methionine residues to produce a significant anomalous signal when replaced by seleno-methionine (Hendrickson, 2014). Incorporation of selenium into the recombinant protein can be done using auxotrophic bacterial strains and introducing seleno-methionine into the growth media during protein expression. Isomorphism between native and seleno-methionine derivative proteins may also be important in highlighting the biological significance of our structure. That is, we need not worry about structural changes affecting our interpretation of the model.

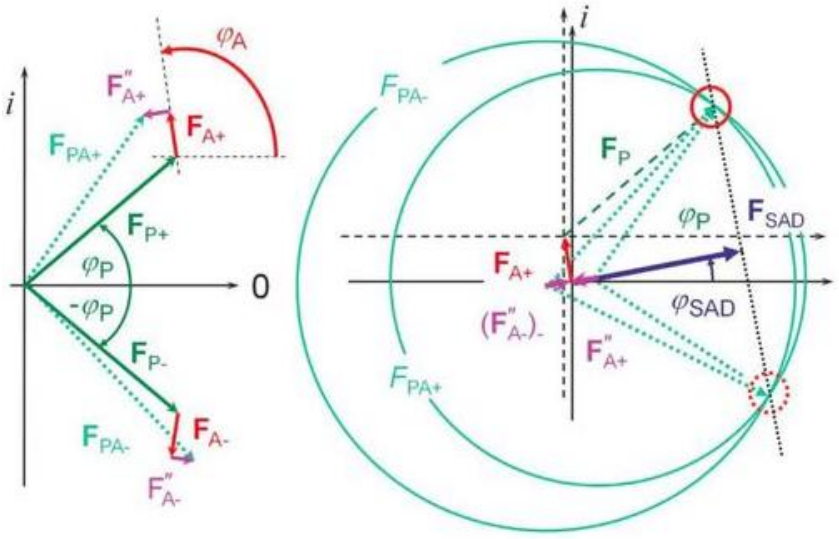

Figure 4. 2 Argand (left) and Harker (right) diagrams depicting phase ambiguity arising from the single anomalous dispersion (SAD) method. The structure factor F_P+ and its Friedel mate, F_P- are shown in green. The anomalous contributions are shown in red (*f'*) and magenta (*f"*). By inverting $F_{PA}-$ in the real axis, we see two points of intersection for structure factors of magnitude $F_{PA}-$ (larger circle) and $F_{PA}+$ (smaller circle). These circles are centred at the head of each unique *f"* contribution. The phase angle selected for SAD is the average of these potential solutions (blue arrow).

4.7.6. Model building and refinement

Electron density maps are visualized using molecular graphics software such as COOT. With high resolution data and knowledge of the protein sequence, initial models can often be built automatically. These models require limited visual inspection and correction by the user. At lower resolution, user experience is critical to model building, as molecular features are difficult to discern. In either case, model building is followed by computational refinement of molecular parameters (stereochemistry, bond lengths, etc.). This process of manual building and refinement is repeated until further improvements to the structure can no longer be made (Salunke and Nair, 2017).

The refinement software scores model quality using an R-factor (R_{work}), which quantifies the agreement between experimental structure factors and those calculated from the model (equation 4.5).

$$R_{work} = \sum_{hkl} [|F_{obs}(hkl)| - |F_{calc}(hkl)|] \,/\, \sum_{hkl} F_{obs}(hkl) \qquad 4.5$$

Since both the model and experimental data are used in the calculation of electron density, refinement can be biased in favor of the model. Another R-factor (R_{free}) was introduced to overcome the problem of model bias. R_{free} indicates how well the structure factors calculated from a model can predict the 5 – 10% of reflections that have been excluded from the refinement process (Salunke and Nair, 2017).

Two maps are often used to guide model building: $2F_o - F_c$ and $F_o - F_c$, where F_o are the structure factors obtained experimentally (observed) and F_c are the corresponding structure factors calculated from the model. The $2F_o - F_c$ map provides an outline of molecular features, while $F_o - F_c$ shows areas of excessive or insufficient electron density (Salunke and Nair, 2017; Wlodawer et al., 2008).

4.8. List of Materials

4.8.1. List of primers used in thesis projects

4.8.1.1. LpnE/OCRL

Forward LIC sequence	5’-TACTTCCAATCCAATGCC-3’
Reverse LIC sequence	5’-TTATCCACTTCCAATGTTA-3’
LpnE-22 forward	5’-GCTACGACTCAGGAAGGTGAGGAG-3’
LpnE-73 forward	5’-GAAAAAACAGAGCAATTATTACTTGCCTC-3’
LpnE-158 forward	5’-AACTCTAATGCGGCTTTGGCTATTGG-3’
LpnE-176 forward	5’-AAAAAGGATAAGACTCAAGCTCTCAACT-3’
LpnE-375 reverse	5’-TTTTTTGGCCTTGTAATTATCCGTATATTG-3’
OCRL-140 reverse	5’- GTCCTTAGTGTCTAATTTCTGGTACCA-3’
OCRL-176 reverse	5’- TTCCCGATGAATCCCAGTAGGCTGATT-3’
OCRL-183 reverse	5’-TGAAAAGGGTGGAGGTGGGGGTTCCCGATG-3’
OCRL-208 reverse	5’- CTTCCGCATGGTGTTGGTCACTTTGGG-3’
OCRL-10 forward	5’-CCGCTTGCCACTGTCGAGGGTATGGAGATG-3’

4.8.1.2. MavE

MavE-39 forward	5’-ACTAGATTTGAAAGAAATTTCCTGATTAATA-3’
MavE-172 reverse	5’-TTCGTCTTTGAGTTTGGCAATTAATTCTT-3’

4.8.1.3. MavL

MavL-42 forward	5’-AATGGCCTATCAATTGTTGCTCAGTAAAG-3’
MavL-388 reverse	5’- ATTAAATACAATGCCATTTTCTCTGACATAATC-3’
MavL-435 reverse	5’-ACTATGAGAAAAAAGACTGAA-3’
Ube2q1-1 forward	5’-CAGCAGCCGCAGCCGCAGGGGCAGCAGCAG-3’
Ube2q1-422 reverse	5’-GCCGTCTTCTTTTGGGGGTGTGTACCAGCC-3’

4.8.2. List of chemicals

100 mM dCTP and 100 mM dGTP	NEB, N0446S
100×BSA	NEB
30% Acrylamide/Bisarcylamide(29:1)	Bioshop, ACR009.500
Adenosine diphosphate ribose	Sigma Aldrich, AO752-25MG
Ammonium persulfate (APS)	BioShop, AMP001.25
Coomassie Brilliant Blue R-250	AMRESCO, (0427-25G)
Dithiothreitol (DTT)	Thermo Fisher Scientific, BP17225
DNA loading buffer	Norgen, 28129
DpnI	NEB, R0176L
Enrich SEC650	Bio-Rad, 780-1650
Enrich SEC70	Bio-Rad, 780-1070
Gelgreen nucleic acid gel stain	Biotium, 41005
Glycerol	Thermo Fisher Scientific, G31-4
Glycine	Thermo Fisher Scientific, BP381-5
HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid)	BioShop, HEP001.1

HiTrap chelating column	GE Healthcare Life Science, 17-0409-03
Imidazole	BioShop, IMD508.1
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Thermo Fisher Scientific, BP1755100
KOD hot start DNA polymerase kit	Novagen, 71086
MPD (Methyl-pentane-diol)	Hampton Research, Aliso Viejo, CA
Ni^{2+} (II) sulfate hexahydrate	Sigma-Aldrich, 227676-500G
Ni-NTA agarose	Qiagen, 142338540
Poly-ethylene glycol (PEG) Molecular 8000	Hampton Research, Aliso Viejo, CA
Potassium phosphate dibasic (K2HPO4)	BioShop, PPD303.1
Potassium phosphate monobasic (KH2PO4)	Thermo Fisher Scientific, P285-500
Protein marker	Bio-Rad, 161-0377
QIAprep Spin Miniprep Kit (250)	Qiagen, 27106
QIAquick Gel Extraction Kit (250)	Qiagen, 28706
Sodium chloride (NaCl)	BioShop, SOD002.10
Sodium dodecyl sulfate (SDS)	BioShop, SDS001.1
Superdex 200 10/300	GE Healthcare Life Science, 28990944
Tris-HCl	BioShop, TRS002.1
T4 DNA ligase and corresponding buffer	NEB, M0202S
T4 DNA polymerase and corresponding buffer	NEB, M0203L
TEMED (N,N,N′,N′-Tetramethylethane)	BioShop , TEM001.25
Tryptone	Thermo Fisher Scientific, BP1421-2
Yeast Extract	BioShop, YEX401.205

5. Structure of the *Legionella* Effector LpnE and its interaction with OCRL

5.1. Literature Review on LpnE and OCRL

5.1.1. LpnE is an effector protein that plays a role in host cell invasion

LpnE (*Legionella pneumophila* Entry) was identified by Newton and colleagues in a study seeking to explain the difference in virulence between *L. pneumophila* and *L. micdadei* (Newton et al., 2006). Specifically, the authors intended to understand the increased virulence of *L. pneumophila* relative to other *Legionella* species. *L. pneumophila* accounts for ~90% of Legionnaire's disease in North America (Benson and Fields, 1998; Yu et al., 2002) and ~50% in Australia (Yu et al., 2002), whereas *L. micdadei* and *L. longbeachae* were found responsible for only 5% of cases worldwide (Benin et al., 2002; Muder and Yu, 2002). They used genomic subtractive hybridization to reveal genetic differences between *L. pneumophila* and *L. micdadei*. Genes found only in *L. pneumophila* were investigated for their ability to confer virulence. One among these was *lpg2222*, which encodes an effector with eight predicted tetratricopeptide (TRP) repeats. This protein was suspected to play a role in host cell invasion because of its similarity to another *L. pneumophila* TRP effector called EnhC, which shares this function. Further studies showed *L. pneumophila* lacking *lpg2222* to be less efficient at invading THP-1 and A549 cell lines than wild-type bacteria. Importantly, neither intracellular replication nor host cell attachment were affected in *lpg2222*-deficient *L. pneumophila*. Based on these findings, the TRP protein encoded by *lpg2222* was named LpnE (*Legionella pneumophila* Entry) for its role in promoting host cell entry (Newton et al., 2006).

5.1.2. Structure and Function of tetratricopeptide (TRP) repeat proteins

Tetratricopeptide (TRP) repeats are 34-residue structural motifs comprised of a pair of antiparallel α-helices. These repeats occur in tandem arrays, which fold into a right-handed super-helix with a concave amphipathic surface implicated in protein-protein interactions

(Figure 5.1). Inter- and intra-repeat stabilization is achieved by hydrophobic interactions between consensus residues (Cerveny et al., 2013; Urosev et al., 2013) and, in rarer cases, with disulfide bonds (Lüthy et al., 2002, 2004). Indeed, individual repeats are composed of small and large hydrophobic residues that form the scaffold upon which ligand binding residues are grafted (Cerveny et al., 2013). TRP repeat-containing proteins participate in various cellular processes including transcription, kinase inhibition, protein folding and immunity (Mittl and Schneider-Brachert, 2007). This variability in function arises from binding promiscuity at the concave surface of the TRP super-helix. Although the source of promiscuity remains unclear, it is likely a consequence of non-covalent interactions mediated by uncharged polar side chains (Cerveny et al., 2013). A prospective binding partner could hydrogen bond to these residues along the inner surface of the super-helix. This would allow the TRP protein to clasp around its partner without employing a specific binding cleft (Figure 5.1).

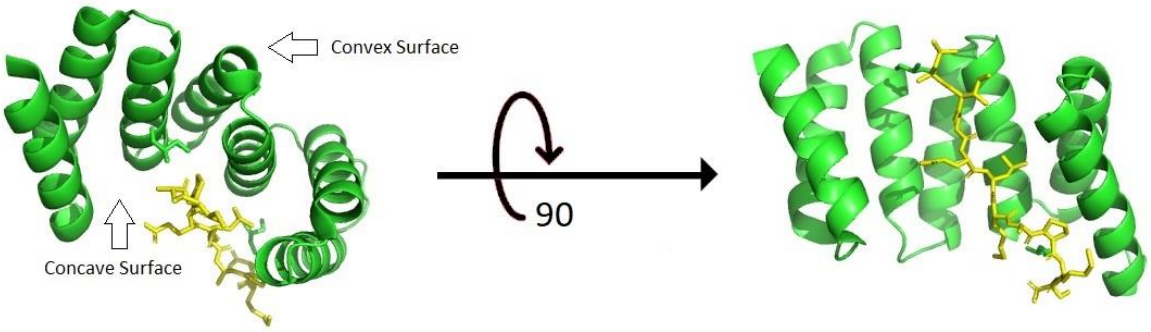

Figure 5. 1 Crystal structure of the tetratricopeptide (TRP) repeat protein IpaC (green) in complex with the IpaB(51-72) peptide (yellow). (Left) View of the complex from above, depicting the outer convex and inner concave surfaces of the TRP super-helix. (Right) Side view of the complex, showing the IpaB(51-72) peptide to adopt an extended conformation in the concavity. Figure adapted from (PDB id 3GZ1).

The order in which TRP repeats fold is governed by the relative stability of each repeat. That is, the most stable repeats fold first, followed by neighboring repeats in a cooperative process. Similarly, when a single TRP repeat unfolds, adjacent repeats do so as well. On this basis, terminal repeats are more likely to unfold than others because they have only one

neighboring repeat (Perez-Riba et al., 2018). There are exceptions to this, since intrinsic repeat stabilities must also be considered in the folding and unfolding processes. In fact, cooperativity of folding depends on the mismatch between intrinsic and interfacial repeat stabilities. Thus, a central TRP repeat may be sufficiently unstable to initiate unfolding, despite being steadied by two neighboring repeats (Perez-Riba et al., 2018).

The cooperative folding of repeat proteins may be linked to their function. Indeed, two 'banana-shaped' repeat proteins acting as allosteric modulators have been described: the HEAT repeat protein PR65 and the LRR (leucine-rich repeat) protein Skp2 (Figure 5.2). These proteins both promote enzymatic activity of their substrates by binding a recognition subunit at one end and a catalytic subunit at the other. Importantly, this pattern of allosteric behaviour was found to depend on the overall shape of the repeat protein and not the repeat identity (Perez-Riba et al., 2018). This may implicate the cooperative folding pattern of these proteins in their allosteric properties. For example, protein-protein interactions occurring at the N-terminus of a repeat protein could lead to conformational changes that are transmitted to the C-terminus in a cooperative fashion. These studies show TRP repeat proteins to be dynamic structures that bridge molecular events by transmitting a conformational signal across the repeats.

The best studied TRP proteins in bacterial pathogens are class II chaperones, which belong to the type III secretion system and deliver two hydrophobic translocators into the eukaryotic membrane. Together with a third hydrophilic stabilizing protein, the two translocators form a pore in the eukaryotic membrane called a translocon (Cerveny et al., 2013). A crystal structure of the *Pseudomonas aeruginosa* class II chaperone PcrH (TRP) in complex with a peptide belonging to the minor translocator PopD showed these proteins to interact along the concave surface of the super-helix (Job et al., 2010). The binding interface is also employed by the major translocator (PopB) when interacting with PcrH (Discola et al., 2014). Thus, major (PopB) and minor (PopD) translocators are each bound independently to the concave surface of PcrH. This finding demonstrates how binding promiscuity of the TRP super-helix is utilized in the delivery of translocon subunits to the eukaryotic membrane.

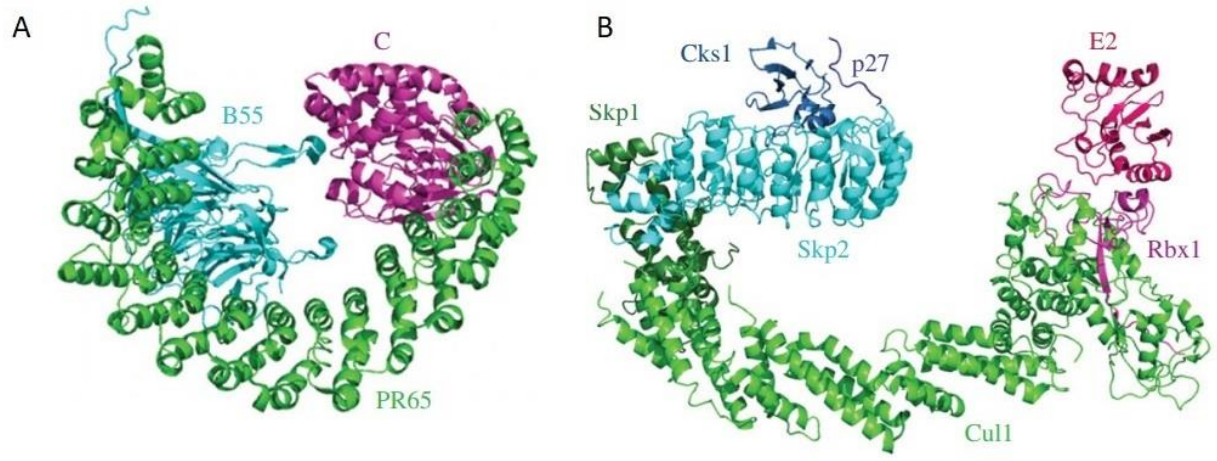

Figure 5. 2 Conformational dynamics of tetratricopeptide repeat (TRP) proteins may transmit molecular signals from one end of the super-helix to the other. (A) HEAT-repeat protein PR65 (green) binding a catalytic subunit (C) at the C-terminus and a regulatory subunit (B55) at the N-terminus. (B) Leucine-rich repeat (LRR) protein Skp2 forms a hinge-like interaction with Skp1. Cks1 binding causes tightening of this hinge, leading to reduced mobility of Skp2. These conformational changes may be transmitted through the Cullin subunit (Cul1) to the E2 ubiquitin-conjugating enzyme.

The role of TRP proteins in translocon formation may point to a similar role for LpnE/EnhC as chaperones in host cell entry. For example, by binding proteins that locally disrupt eukaryotic membrane integrity, LpnE may facilitate the entry of an entire bacterium into the host cell. This suggestion is consistent with the observed role of TRP proteins as adaptors in the formation of macromolecular complexes (Cerveny et al., 2013; Mittl and Schneider-Brachert, 2007). Further studies are required to identify binding partners of LpnE relevant to its role in entry. Another factor that warrants consideration is the dynamic properties of TRP proteins. This adds an additional layer of complexity to deciphering the function of TRP proteins; binding promiscuity and cooperative conformational changes transmitted through the repeats may lead to multiple functions for a single TRP protein.

Sel1-like repeat (SLRs) are a subclass of the TRP family, distinguished by their longer length and greater variability in the consensus sequence. Instead of 34-residue motifs, SLRs contain 36 – 44 residues. This difference in length arises from a 4-12 residue insertion in one loop and a 2-residue deletion in the other. Superposition shows the SLR super-helix to be wider than that of TRPs due to these loop-length variations (Figure 5.3) (Mittl and Schneider-

Brachert, 2007). Both LpnE and EnhC belong to the SLR class of proteins and harbor 9- and 21-repeats, respectively. The fact that LpnE promotes host cell entry (Newton et al., 2006) may suggest that EnhC does not require its full SLR complement to do the same. Alternatively, these proteins may facilitate host cell entry by different methods employing unique adaptor or pore-forming proteins.

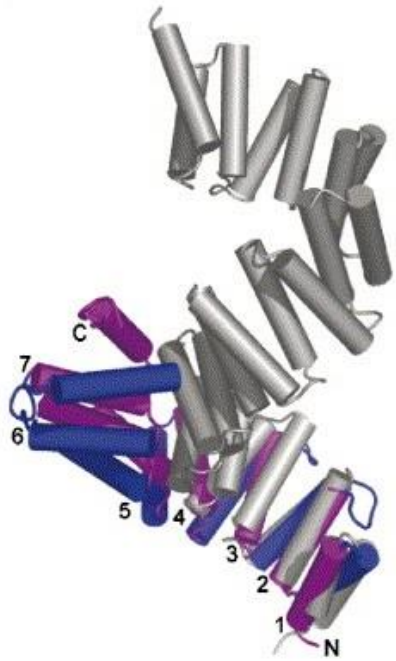

Figure 5. 3 Superposition of the SLR protein HcpC onto the TRP domain of O-linked GlcNac transferase. Helices 1 and 2 of HcpC are shown in magenta and blue, respectively, while the GlcNac transferase is shown in grey. Note the tighter turns in the TRP repeat, relative to the SLRs. Figure adapted from (Mittl and Schneider-Brachert, 2007).

5.1.3. Additional roles of LpnE in *Legionella pneumophila* pathogenesis

LpnE promotes the infection and intracellular replication of *L. pneumophila* in *Acanthamoeba castellanii* and A/J mice using a full complement of SLRs, as deletion of only two C-terminal repeats rendered an invasion defect in *L. pneumophila*. Once inside the cell, *L. pneumophila* uses LpnE to modulate LCV trafficking. This was evidenced in a study where *L. pneumophila* lacking *lpg2222* produced LCVs with an impaired ability to avoid vesicles positive for lysosome-associated membrane protein-1 (LAMP1) (Newton et al., 2007b). Thus,

LpnE appears to direct the LCV away from canonical endocytic trafficking routes leading to lysosomal degradation. Finally, LpnE is exported from bacterial cells in the absence of both the Icm/Dot type IV secretion system and Lsp type II secretion system (Newton et al., 2007b). The mechanism by which LpnE is exported from *L. pneumophila* without these secretion systems remains unclear.

5.1.4. Oculocerebrorenal syndrome of Lowe (OCRL) protein

Oculocerebrorenal syndrome of Lowe (OCRL) is an inositol polyphosphate 5-phosphatase that regulates endocytosis, autophagy and retrograde trafficking in eukaryotic cells. The regulatory effect of OCRL arises from its phosphatase activity, which alters vesicular lipid composition (De Matteis et al., 2017) by dephosphorylating PI(4,5)P2, Ins(1,4,5)P3 and PI(3,4,5)P3. The reduction of PI(4,5)P2 by OCRL on early endosomes facilitates their progression along the endocytic pathway. Dephosphorylation of PI(4,5)P2 to PI4P causes the dissociation of clathrin from early endosomes, because PI(4,5)P2 anchors clathrin to membranes. OCRL also regulates endo-lysosomal fusion under conditions of sustained autophagic flux, where PI 5-kinase enzymes increase the presence of PI(4,5)P2 on the lysosome. Uncontrolled accumulation of PI(4,5)P2 limits the fusion of autophagosomes with lysosomes by inhibiting the TRPML1 calcium channel. Thus, OCRL maintains PI(4,5)P2 at appropriate levels on the lysosome under these conditions (Dong et al., 2010; Zhang et al., 2012). Finally, overexpression of OCRL in HeLa cells impedes retrograde trafficking of cation independent mannose 6-phosphate receptor (CI-MPR) and clathrin. This likely arises from the association of OCRL with the *trans*-Golgi network (TGN) through Rab1 or Rab6, since CI-MPR and clathrin are redistributed from the TGN to late endosomes (Dressman et al., 2000; Olivos-Glander et al., 1995). These findings consistently show that OCRL requires its phosphatase activity for regulating vesicular trafficking events.

Mutations in the oculocerebrorenal syndrome of Lowe (OCRL) protein give rise to a disease of the same name (Lowe syndrome) characterized by dysfunction of proximal renal tubules (Zaniew et al., 2018). These tubules are lined with epithelial cells that take up proteins by clathrin-mediated endocytosis, which almost eliminates plasma proteins from human urine

under physiological conditions. A brief overview of this process is given below to provide context for the role of OCRL in Lowe syndrome.

Clathrin-mediated endocytosis starts with the interaction of two multiligand receptors, megalin and cubulin, which are anchored to the membrane by another protein called amnionless (AMN). Once ligand binds to cubulin, the megalin/cubulin/AMN complex interacts with the clathrin coat using adaptor proteins such as disabled homologue 2. Thus, ligand binding induces uptake of the megalin/cubulin/AMN complex into clathrin-coated pits that separate from the membrane to form early endosomes (Coudroy et al., 2005). Sustained acidification of these endosomes by the vacuolar H^+-ATPase leads to ligand dissociation and subsequent uncoating of clathrin (Herak-Kramberger et al., 1998). These uncoated vesicles are termed late endosomes, which fuse with lysosomes for the degradation of unbound ligand by lytic enzymes contained therein.

Insights into how higher eukaryotes combat *L. pneumophila* infection were obtained in a study showing LpnE to bind the eukaryotic protein, Oculocerebrorenal syndrome of Lowe (OCRL). As described in section 2.1.8, OCRL plays a role in retrograde trafficking. When recruited to the LCV by its interaction with LpnE, OCRL restricts intracellular replication of *L. pneumophila* by an unknown mechanism (Weber et al., 2009). This contrasts with the beneficial role of OCRL in enriching PI4P on the LCV outer surface. Also, *L. pneumophila* effectively diminishes the proportion of OCRL available to mediate canonical endocytic trafficking. As such, the recruitment of OCRL to the LCV may initially benefit *L. pneumophila* but, ultimately, this interaction impedes replication.

Importantly, the phosphatase domain of OCRL is not contained within the construct found to interact with LpnE. In fact, only the N-terminal 236 residues of OCRL are recruited to the LCV by LpnE (Weber et al., 2009). This OCRL construct comprises a pleckstrin homology (PH) domain followed by an unstructured linker. The phosphatase domain lies C-terminal to this linker. Therefore, LpnE brings the phosphatase domain of OCRL in proximity to the LCV without directly binding it. In this way, the phosphatase domain can exert its enzymatic activity on the LCV without being affected by protein-protein interactions. It remains unclear if the LpnE/OCRL interaction allows the LCV to avoid endocytic trafficking, but it would facilitate

the production of PI4P on the LCV at early stages of infection. Further studies are needed to clarify the role of OCRL in LCV trafficking.

5.2. Experimental Procedures for the LpnE and OCRL project

5.2.1. Cloning of recombinant LpnE

The entire lpnE (*lpg2222*) gene was amplified from *Legionella pneumophila* (Philadelphia) genomic DNA by PCR. Since residues 1–22 of LpnE were predicted to encode a signal peptide (SignalP, http://www.cbs.dtu.dk/services/SignalP/), the DNA sequence encoding residues 22–375 was also amplified. These constructs were designed using the ligation-independent cloning (LIC) method. Digesting LpnE(22-375) with trypsin led to the production of a stable fragment, which was identified by mass spectrometry to be LpnE(73-375). This construct was also cloned into pMCSG7 by the LIC method. Finally, two truncated LpnE constructs were also designed: LpnE(158-375), LpnE(176-375) by this approach.

5.2.2. Cloning of recombinant OCRL

A plasmid containing the human OCRL gene was purchased from Addgene (pcDNA3-HA-human OCRL, plasmid #22207). The following constructs of OCRL were designed: OCRL(1-236), OCRL(10-236), OCRL(1-140), OCRL(10-140), OCRL(10-176), OCRL(10-208). Amplicons were placed in pMCSG7 and a pGEX derivative with a TEV cleavage site (pRL652) via ligation-independent cloning (LIC), as described above for LpnE.

5.2.3. Protein expression and purification

Initially, LpnE(1-375) and LpnE(22-375) were expressed in BL21(DE3) and purified as described below. LpnE(1-375) did not express from pMCSG7 in our hands and exhibited poor solubility as a GST fusion protein. His-LpnE(22-375) expressed well and was eluted from Ni-NTA in 100 mM imidazole and then loaded onto size-exclusion column (SEC650,

Biorad/S200, GE Life Sciences, Mississauga, Canada) for separation of oligomeric states. Analysis of the eluted fractions by dynamic light scattering (DLS) showed LpnE(22-375) to have a high degree of polydispersity.

Since LpnE(22-375) formed various aggregates/oligomers in solution, we investigated if the presence of OCRL(10-208) would improve the behavior of this protein. Cell pellets containing His-LpnE(22-375) and His-OCRL(10-208) were mixed, co-lysed, and purified on Ni-NTA followed by size-exclusion chromatography. This protocol resulted in a single narrow peak containing both proteins. An additional strategy to identify a fragment of LpnE(22-375) with good solubility was limited proteolysis. Digestion with trypsin produced a stable and soluble fragment, which was identified by mass spectrometry to consist of residues 73–375. This trypsin digestion product gave a symmetrical peak on gel filtration and led to the initial crystal hits. Subsequently, LpnE(73-375) was cloned to replicate the trypsin digestion product and used in all further structural experiments.

LpnE(73-375), LpnE(22-375), LpnE(158-375), LpnE(176-375), OCRL(10-140), OCRL(10-176), OCRL(10-208), and OCRL(10-236) plasmids were transformed into the pMCSG7 vector, expressed and purified as described in sections 4.1 – 4.6. Size exclusion chromatography was carried out using a Biorad SEC650 or GE Superdex200 column. The His_6-tags were cleaved after affinity chromatography, overnight at 20 °C using TEV protease at a 1:50 (enzyme:target) ratio.

5.2.4. Crystallization of LpnE(73-375)

Analysis of the LpnE sequence indicates the presence of a signal peptide encompassing residues 1–21. We have cloned and expressed the expected mature form of LpnE containing residues 22–375 in *E. coli*. Although we were able to crystallize LpnE(22-375), these crystals diffracted poorly. We applied limited proteolysis to identify a better candidate for crystallization. A limited digest of LpnE(22-375) showed the presence of a stable 36 kDa fragment. This fragment was identified by mass spectrometry to comprise residues 73–375 and contain most of the predicted SLRs. Subsequently, this fragment was cloned with an N-terminal His_6-tag, overexpressed in *E. coli* BL21(DE3)pLysS cells and used for structural and binding studies.

Monomeric and dimeric fractions of LpnE(73-375) were screened for crystallization by the sitting drop method in 96-well plates using in-house and commercial screens. Only monomeric fractions of LpnE(73-375) showed low polydispersity by dynamic light scattering (DynaPro Plate Reader II, Wyatt Technology, Santa Barbara, CA, USA). The proteins were concentrated to ~60 mg/mL in buffer containing 15 mM Tris-HCl, pH 8.0, and 50 mM NaCl, and preparations with and without the His-tag were screened. Untagged LpnE(73-375) gave extremely fine-needle clusters under a few conditions and optimization failed to improve the crystal size. Conversely, His-tagged LpnE produced single crystals in many different conditions. After optimization by the hanging drop vapor diffusion method, the best crystals were obtained at 20 °C in drops containing 1 µL protein mixed with 1 µL of reservoir solution [20% PEG 8000, 0.2 M ammonium sulfate, 0.1 M 2-(N-morpholino)ethanesulfonicacid (MES), pH 6.5] and suspended over 500 µL reservoir solution. The LpnE(22-375)-OCRL(10-208) complex displayed low polydispersity and was screened with the His-tag and with the tag removed by TEV protease. Only His-tagged preparations gave crystals after 5–7 days of growth at 20 °C under various conditions within the ProComplex Suite. Examination of the crystal content by SDS/PAGE revealed that His-LpnE(22-375) was the sole constituent. These crystals diffracted poorly and were recalcitrant to optimization.

5.2.5. Data collection and structure determination

The His-LpnE(73-375) protein crystals were cryoprotected by transferring to 1 µL mother liquor containing 20% (v/v) glycerol, 0.2 M ammonium sulfate, 20% PEG8000, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) pH 6.5. Diffraction data were collected to 1.75 Å resolution at the Canadian Macromolecular Crystallography Facility (CMCF) 08ID beamline, Canadian Light Source, using a Rayonix MX300_CCD Detector (Grochulski et al., 2011). Integration and scaling were carried out using the XDS software package (Kabsch, 2010) (Table 1). The structure was solved by molecular replacement using the MOLREP program (Vagin and Teplyakov, 2010). Residues 28–285 of *H. pylori* cysteine-rich protein C (HcpC, PDB code 1OUV, (Lüthy et al., 2004)) were used as a search model, and two molecules were expected in the asymmetric unit based on Matthew's coefficient. This molecular replacement solution showed good fit to the electron density only for the N-terminal half of the protein,

indicating a slightly different relative orientation of the N- and C-terminal segments between LpnE and the model structure. Therefore, the model protein was divided into two parts and used independently for molecular replacement with Phaser (McCoy, 2017). The final solution showed a small reorientation of the two fragments. This model was rebuilt with phenix.autobuild script and refined using phenix.refine (Adams et al., 2011) with intermittent manual rebuilding with Coot (Emsley et al., 2010). The coordinates and structure factors have been deposited with the Protein Data Bank with ID code 6DEH.

5.2.6. Interaction studies using size exclusion chromatography

The interaction of LpnE(73-375), LpnE(158-375), and LpnE(176-375) with OCRL(10-140), OCRL(10-176), OCRL(10-183), and OCRL(10-208) was evaluated by size-exclusion chromatography. Recombinant proteins were purified as described previously and their hexa-histidine tags were cleaved using TEV protease. Complete digestion of the tag was monitored by SDS/PAGE. Purified and cleaved proteins were either injected onto a Biorad SEC70 column for analysis or mixed with a putative-binding partner at an approximate 1:1 molar ratio. Putative complexes were allowed a minimum of 30 minutes at 4 °C to form before being injected onto the SEC70 column.

5.2.7. Size exclusion chromatography with multi-angle light scattering

The binding of LpnE(73-375) and LpnE(22-375) to OCRL(10-208) and OCRL(10-236) was evaluated using size-exclusion chromatography (SEC) with multi-angle light scattering (SEC-MALS) at room temperature. A 500 µL sample of size-exclusion purified (Biorad SEC650) LpnE and OCRL was injected at 2 mg/mL on an analytical size exclusion column, S200 Increase (Wyatt Technologies, Santa Barbara, CA), using AKTA Explorer FPLC system (GE Healthcare). Elution from the column was passed on to the MALS system comprising MiniDawn TREOS (Wyatt Technology, Santa Barbara, CA). The column was equilibrated at room temperature with buffer comprised of 15 mM TrisHCl, pH 8.0, and 50 mM NaCl. Detector normalization was achieved using 2 mg/mL BSA (Pierce, Burlington, Canada).

5.2.8. Isothermal titration calorimetry of LpnE(73-375) with OCRL(10-208)

Titrations were carried out using the Nano ITC instrument (TA Instruments, New Castle, DE). 200 µM OCRL(10-208) in a buffer containing 15 mM Tris-HCl, pH 8.0, and 50 mM NaCl was titrated into the calorimeter cell containing 50 µM LpnE(73-375) in the same buffer. Experiments were performed at 20 °C on untagged proteins. Data analysis was done with NANOANALYZE software (TA Instruments, New Castle, DE, USA) using an independent binding model (see section 4.5.2 for detailed explanation of ITC).

5.2.9. Cloning LpnE for localization in human cells

To obtain the GFP fusion expression vector, LpnE(1-375) and LpnE(73-375) were PCR-amplified from the genomic DNA of *L. pneumophila* (strain Philadelphia). Fragments were digested with XhoI and BamHI and ligated to the same restriction sites of pEGFP-N1 (Clontech Laboratories, Mountain View, CA, USA) expression vector. Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (NEB, Ipswich, MA, USA). To obtain LpnE(22-375)-GFP, the signal peptide was looped out of pEGFP-N1 harboring LpnE(1-375). Successful cloning of each plasmid was confirmed by DNA sequencing.

5.2.10. Transient transfection of HEK293 cells

Human Embryonic Kidney cell line 293 (HEK293) was cultured in Dulbecco's Modified Eagle Medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% Fetal Bovine Serum (FBS) (Sigma-Aldrich) at 37°C with 5% CO_2. DNA constructs were transfected into HEK293 cells using the X-treme GENE^{TM} HP DNA Transfection Reagent (Roche, Cat. 06366236001, Mississauga, Canada) according to the manufacturer's instructions.

5.2.11. Immunofluorescence

GFP-tagged LpnE(1-375), LpnE(22-375), or LpnE(73-375) expressing cells were grown on 12 mm diameter glass cover-slips. Cells were fixed with 4% paraformaldehyde solution prepared in phosphate buffered saline (PBS) for 30 min at 20°C, washed in PBS, permeabilized with 0.5% Triton X-100 in PBST, and blocked in 5% normal horse serum for 20 min. Cells were then incubated with primary anti-GM130 antibody (BD Transduction Laborato-ries, Mississauga, Canada) for 60 min at room temperature at a dilution of 1:200 in the blocking solution. Cells were washed with PBST, and a secondary Alexa Fluor 546 goat anti-mouse IgG (Invitro-gen, Carlsbad, CA, USA) was overlaid on coverslips for 20 min at 20°C at a dilution of 1:2000 in blocking solution. Slides were mounted and visualized on a Laser Scanner Confocal Microscope (Zeiss LSM700, Thornwood, NY, USA).

5.3. Results

5.3.1. LpnE/OCRL Binding Studies

LpnE has been shown by immunoprecipitation to interact with the N-terminal region of OCRL comprising residues 1–236 (Weber et al., 2009). This region encompasses a pleckstrin homology (PH) domain (aa 1–119) (Mao et al., 2009) and a long flexible linker connecting the PH domain to the inositol polyphosphate 5-phophatase catalytic module (Trésaugues et al., 2014). Although PH domains have been implicated in protein–protein interactions, we suspected that the unstructured linker of OCRL may harbor binding elements essential for its interaction with LpnE.

His-LpnE(73-375) eluted from the SEC column in two peaks corresponding to monomeric and dimeric species. The monomer was homogeneous, and we first explored if this fragment binds OCRL(1-236). We have shown using size-exclusion chromatography (SEC) that LpnE(73-375) does indeed interact with GST-OCRL(1-236), as evidenced by a distinct peak shift for the His-LpnE(73-375)–GST-OCRL(1-236) complex relative to His-LpnE(73-375) and GSTOCRL(1-236) alone (Figure 5.4A). Secondary structure predictions (PsiPred, [21]) of

OCRL show that the region 135–200 neighboring the PH domain is devoid of secondary structure and is followed by a short helix (aa 201-206). A polyproline motif (PPPPP) is present within this unstructured region (aa 177-181) and may also promote the binding of OCRL to LpnE [22]. Moreover, the first nine N-terminal residues of OCRL are a string of hydrophobic residues. To define which part of OCRL(1-236) is crucial for interacting with LpnE, we designed the following constructs of OCRL based on the above information: OCRL(1-140) and OCRL(10-140) that correspond to the PH domain, OCRL(10-176), OCRL(10-183), OCRL(10-208), and OCRL(10-236). These constructs were cloned with a cleavable His_6-tag. All recombinant proteins were purified by affinity chromatography using Co^{2+} Talon resin and eluted with a buffer containing 100 mM imidazole. The His_6-tag was cleaved with TEV protease overnight at room temperature. All OCRL constructs, except the shortest OCRL(10-140), were found to have a degradation product corresponding to ~15 kDa after TEV cleavage. We found the longest construct, OCRL(10-236), to be poorly expressed and vulnerable to proteolytic degradation, making data interpretation difficult for this protein. We therefore concentrated our efforts on the other constructs. One of these, OCRL(1-140), includes the first nine residues to observe their role in binding LpnE. Our C-terminal truncations of OCRL systemically probe into the region without predicted secondary structure and should provide insight into the residues required to promote an interaction with LpnE.

The interaction of untagged LpnE and OCRL constructs was investigated by size-exclusion chromatography using a BioRad SEC70 column. The proteins were mixed in an approximate 1 : 1 molar ratio and incubated for 30 minutes on ice. When LpnE(73-375) was combined with OCRL(10-208), a clear shift in the elution profile was observed, with the main peak migrating to a shorter elution time and containing both proteins. This indicated that the two proteins form a complex (Figure 5.4B). To quantify their binding, we measured the binding constant for LpnE(73-375) with OCRL(10-208) by isothermal titration calorimetry and obtained a value of 16 µM (Figure 5.4C) confirming a modest binding strength. Next, we tested if the shorter OCRL(10-183) interacts with LpnE(73-375). A shift in the profile is also visible for this complex, with fractions containing both proteins appearing at lower elution volumes (Figure 5.4D). Additional deletion of the polyproline segment from OCRL (OCRL(10-176)) leads to a substantial weakening of the interaction with LpnE. This is evidenced by a small shift to

lower volumes and limited overlap of fractions containing the proteins (Figure 5.4E). We interpret the small shift of the peaks as indicating a dynamic formation and dissociation of the complex, which increases somewhat the apparent molecular weight of each protein. Finally, the shortest OCRL constructs, OCRL(1/10-140), do not interact with LpnE(73-375); they show no shift in the elution profile and no overlap of the fractions containing each protein is observed (Figure 5.4F). We conclude that the polyproline region in OCRL significantly contributes to the binding with LpnE.

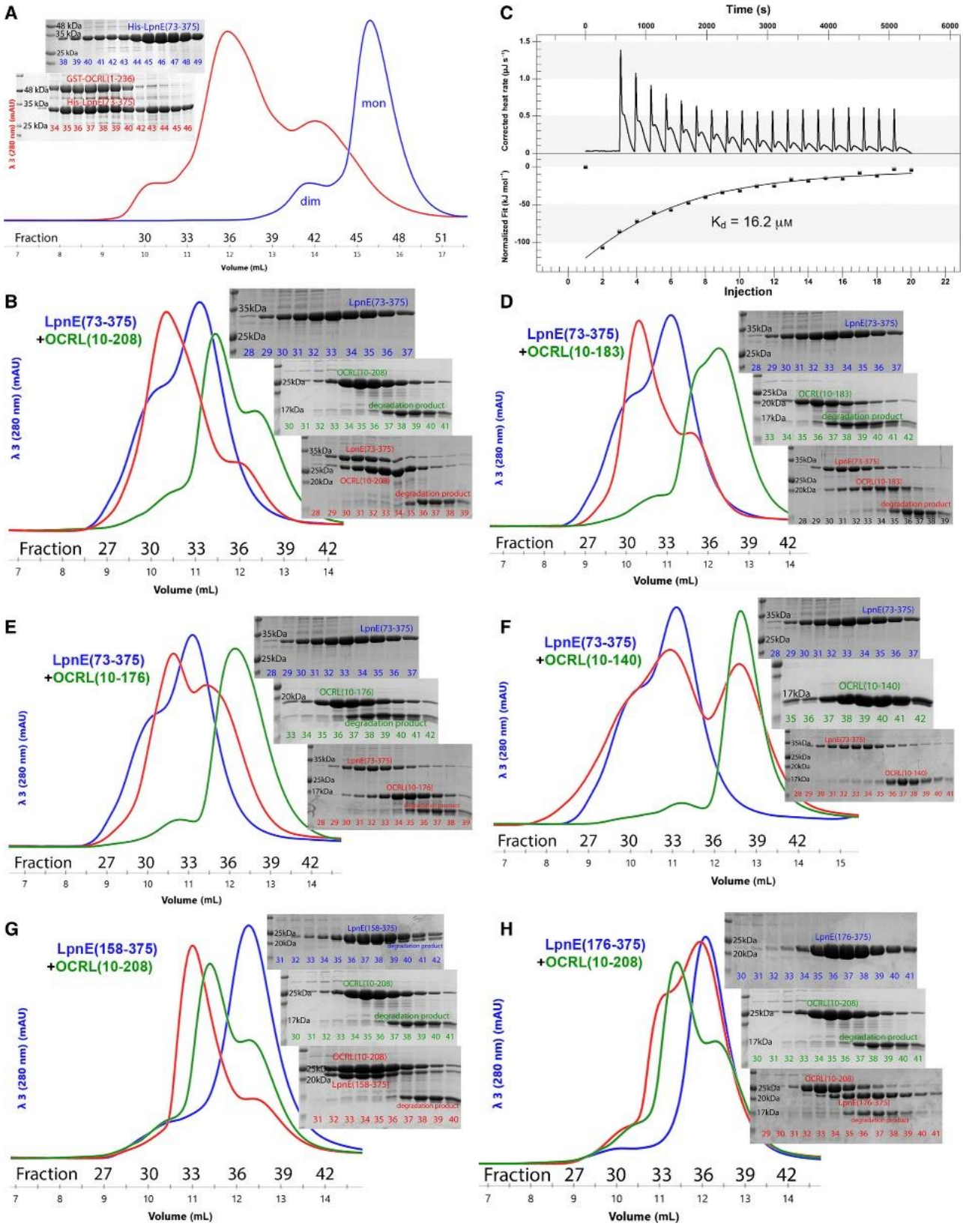
A
His-LpnE(73-375)
GST-OCRL(1-236)
His-LpnE(73-375)
mon
dim
λ 3 (280 nm) (mAU)
Fraction
Volume (mL)
C
Time (s)
Corrected heat rate (μJ s⁻¹)
Normalized Fit (kJ mol⁻¹)
K_d = 16.2 μM
Injection
B
LpnE(73-375)
+OCRL(10-208)
degradation product
D
LpnE(73-375)
+OCRL(10-183)
E
LpnE(73-375)
+OCRL(10-176)
F
LpnE(73-375)
+OCRL(10-140)
G
LpnE(158-375)
+OCRL(10-208)
H
LpnE(176-375)
+OCRL(10-208)

Figure 5. 4 Interaction of LpnE with OCRL fragments.(A) SEC profile of His-LpnE(73-375)–GST-OCRL(1-236) complex and His-LpnE(73-375) alone from BioRad SEC650 column with the SDS/PAGE gel across the protein fraction of the complex. His-LpnE(73-375) elutes in two peaks, monomer (mon) and dimer (dim); (B) Combined SEC profiles and SDS/PAGE corresponding to each profile for LpnE(73-375) in blue, OCRL(10-208) in green and their mixture in red. The proteins were injected on Biorad SEC70 column. The same column was used in all the runs presented this figure and the proteins are marked using the same color scheme; (C) the ITC profile for binding of OCRL(10-208) to LpnE(73-375). Binding constant of 16.2 μM was calculated from the measured data; (D) combined SEC profiles and SDS/PAGE corresponding to each profile for LpnE(73-375), OCRL(10-183), and their mixture; (E) combined SEC profiles and SDS/PAGE corresponding to each profile for LpnE(73-375), OCRL(10-176) and their mixture; (F) combined SEC profiles and SDS/PAGE corresponding to each profile for LpnE(73-375), OCRL(10-140), and their mixture; (G) combined SEC profiles and SDS/PAGE corresponding to each profile for LpnE(158-375), OCRL(10-208), and their mixture; (H) combined SEC profiles and SDS/PAGE corresponding to each profile for LpnE(183-375), OCRL(10-208), and their mixture.

We then explored how far LpnE can be truncated on the N-terminal side before OCRL binding is abolished. To this end, we expressed His_6-tagged LpnE with five- [LpnE(158-375)] or six- [LpnE(176-375)] helices removed from the LpnE(73-375) constructs. These proteins were purified, and their tags were cleaved with TEV protease. We investigated whether OCRL(10-208) retains an ability to bind these truncated LpnE(73-375) constructs. The SEC profile of LpnE(158-375) and OCRL(10-208) shows a shift of the main peak to lower volumes and the presence of both proteins in the peak fractions, indicating retention of binding (Figure 5.4G). This LpnE (158-375) fragment still interacted with OCRL(10-183) (Figure 5.5A) but not with OCRL(10-176) (Figure 5.5B) or OCRL(10-140) (Figure 5.5C). The shorter LpnE construct, LpnE(176-375), mixed with OCRL(10-208) showed almost no shift in peak position (Figure 5.4H), suggesting a much weaker interaction between these two protein fragments. We conclude that the LpnE segment aa 158-177 containing an a-helix plays a key role in OCRL binding.

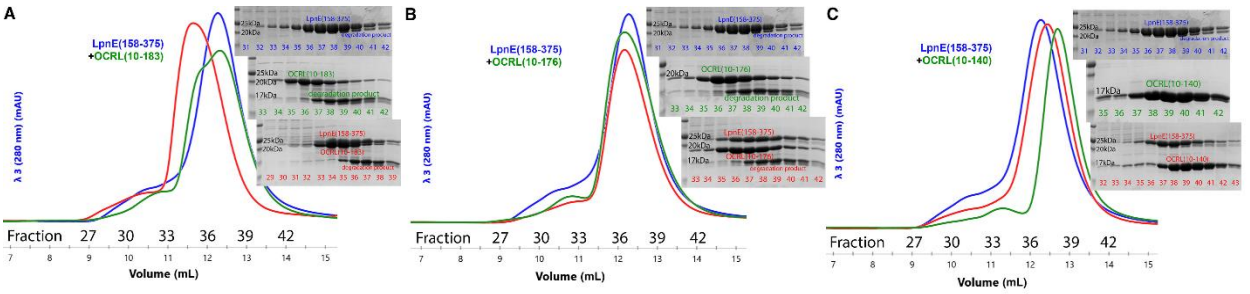

Figure 5. 5 LpnE(158-375)-OCRL interaction profiles. SEC elution profiles of individual constructs and their 1 : 1 mixture and corresponding SDS-PAGE gels of peak fractions are shown. LpnE is in blue, OCRL in green, and the mixture in red. The proteins were injected on the Biorad SEC70 column. (A) LpnE(158-375), OCRL(10-183), and their mixture; (B) LpnE(158-375), OCRL(10-176), and their mixture; (C) LpnE(158-375), OCRL(10-140).

Further analysis of the putative LpnE-OCRL complexes was carried out on untagged proteins using size exclusion chromatography combined with multi-angle light scattering (SEC-MALS). LpnE(22-375) eluted in a peak corresponding to a molecular weight of 41.8 kDa, which indicates that this protein is monomeric (39.4 kDa theoretical) (Figure 5.6A).

Similar to its behavior on the SEC650 column, His-LpnE(73-375) eluted in two distinct peaks representing monomeric (35.7 kDa) and dimeric (71.5 kDa) states of the protein. Cleavage of the hexa-histidine tag abolished LpnE(73-375) dimerization, suggesting a role for the tag in promoting dimer formation (Figure 5.6B). Indeed, untagged LpnE(73-375) gave a molecular weight estimate of 34.5 kDa on SEC-MALS (33.7 kDa theoretical monomer). OCRL(10-208) also proved to be monomeric and eluted in a 23.3 kDa peak on SECMALS (22.9 kDa theoretical monomer) (Figure 5.6C). In agreement with the SEC data, the SEC-MALS interaction of LpnE(73-375) with OCRL(10-208) produced a peak with an accompanying molecular weight estimate of 69.6 kDa, which corresponds loosely to the expected molecular weight for a 1 : 1 complex of these proteins (56.6 kDa theoretical) (Figure 5.6D).

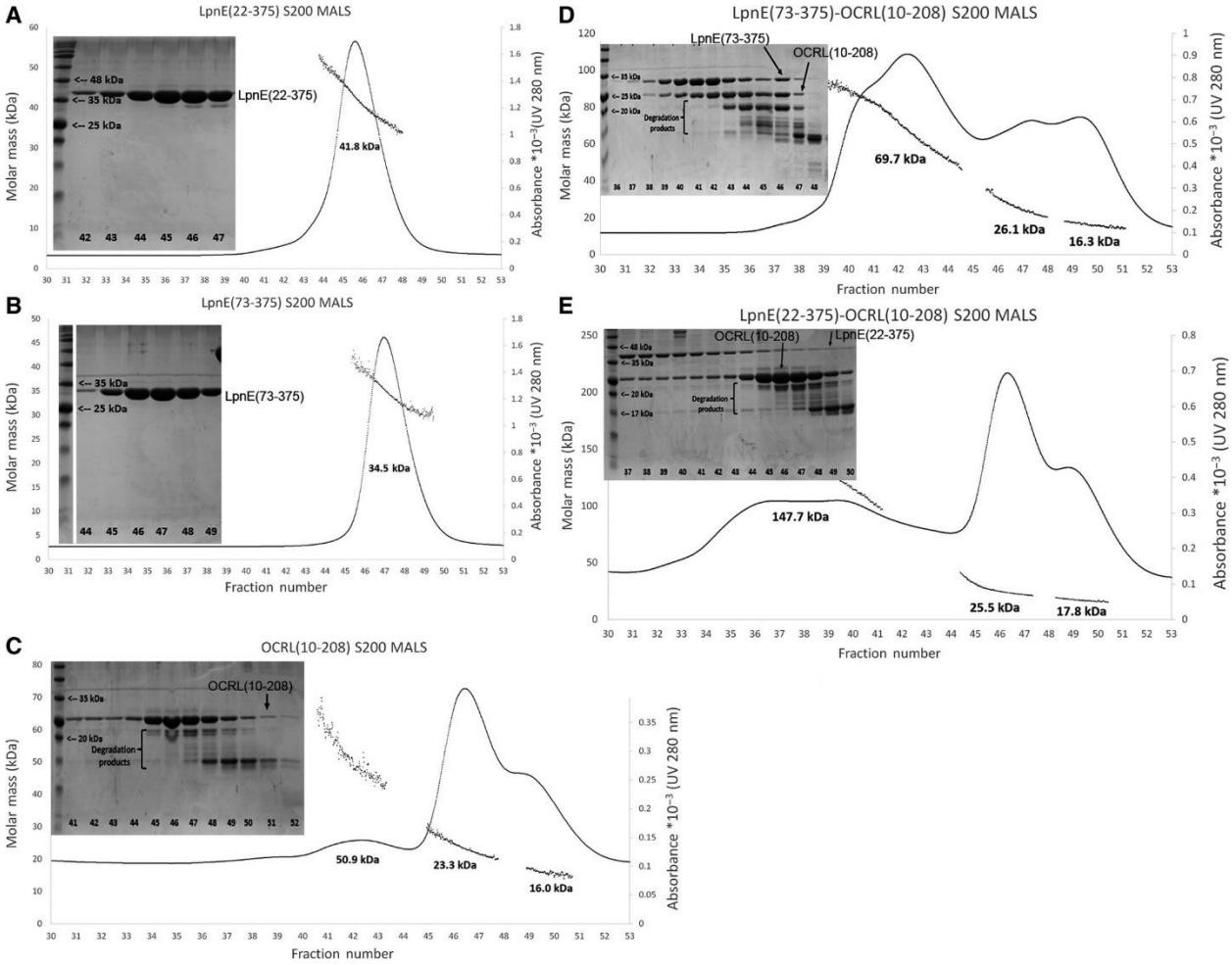

Figure 5. 6 Interaction of LpnE(73-375) or LpnE(22-375) with OCRL(10-208) monitored by SEC-MALS and SDS/PAGE. (A) LpnE(22-375) elutes as a single peak with MW of 41.8 kDa. SDS/PAGE shows the fractions from the column; (B) LpnE(73-375) elutes as a single peak with Mw of 34.5 kDa. The middle part of the SDS/PAGE gel (shown in panel A) has been spliced out to move molecular weight markers close to the relevant lanes of the gel; (C) OCRL(10-208) elutes as a single peak with MW of 23.3 kDa. Degradation products of OCRL elute in the second peak; (D) LpnE(22-375) with OCRL(10-208). The complex elutes within fractions 40–43; (E) LpnE(22-375) with OCRL(10-208). The complex elutes as a broad peak within fractions 36–42. Excess of OCRL(10-208) elutes in the second peak and degradation products as the third peak.

A ~10 kDa deviation from the expected value may suggest that the LpnE(73-375)-OCRL(10-208) complex is not spherical in solution, as MALS molar mass estimates are predicated on this assumption. The LpnE(22-375)- OCRL(10-208) mixture eluted together in a peak with a corresponding molecular weight of 147.7 kDa (Figure 5.6E). This could represent an $LpnE(22\text{-}375)_2$- $OCRL(10\text{-}208)_2$ heterotetramer according to theoretical estimates (130.2

kDa). As for the LpnE(73- 375)-OCRL(10-208) complex, the deviation of LpnE $(22\text{-}375)_2$-$OCRL(10\text{-}208)_2$ from its expected molecular weight may imply the complex is nonglobular.

Intriguingly, conformational homogeneity of full-length LpnE(22-375) was promoted by the presence of OCRL(10-208). When purified alone, LpnE(22-375) produced an extremely broad peak on size exclusion (Figure 5.7, magenta trace) and gave consistently high polydispersity (> 20 %Pd) as measured by dynamic light scattering (DLS). By including an ion exchange step in the purification protocol and increasing the salt concentration in the size exclusion buffer to 600 mM, greater purity of LpnE(22-375) could be achieved, although this came at the cost of significant yield reduction. No crystals could be obtained with the protein prepared by this protocol despite broad screening. When copurified in the presence of OCRL(10-208), however, a simple purification protocol consisting of affinity chromatography followed by ion exchange allowed us to obtain large quantities of the complex at low polydispersity. Unlike LpnE(22-375) alone, the size exclusion peak for the complex was sharp and distinct (Figure 5.7, green trace), suggesting that homogeneity is increased for the complex.

Crystallization trials on the LpnE(22-375)-OCRL(10-208) complex gave hits under several conditions. In particular, the ProComplex Suite (Qiagen) produced crystals in B1, C5, C8, D3, D10, D11, D12, E8, E9, F1, F4 and G5 wells. These crystals would only grow if tags remained intact for both LpnE(22-375) and OCRL(10-208). However, analysis of these crystals by SDS-PAGE showed only LpnE(22-375) to be present within them. The absence of OCRL(10-208) likely reflects a tendency for this protein to degrade, as OCRL(10-208) also disappeared from the stock and drop solutions. Efforts to solve the structure of LpnE(22-375) by crystallography were hampered by the low-resolution data (>5 Å) obtained from these crystals. Consequently, it remains unclear what role OCRL(10-208) plays in facilitating the crystallization of LpnE(22-375).

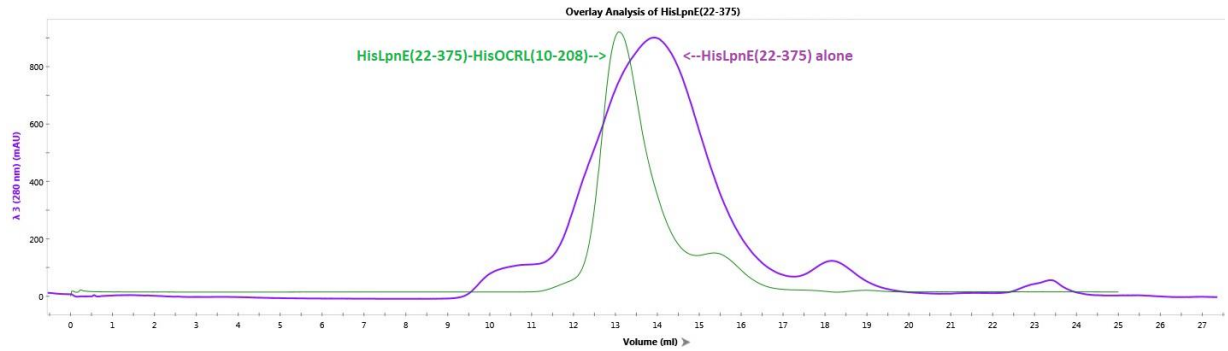

Figure 5. 7 Overlay of size exclusion chromatograms obtained for His-LpnE(22-375) alone (magenta) and the His-LpnE(22-375)-His-OCRL(10-208) complex (green) on a GE Superdex200 column.

Table 5. 1 List of ProComplex Suite Conditions giving rise to crystals of HisLpnE(22-375)

ProComplex Well	**Condition**
B1	0.2 M Lithium sulfate, 0.1 M Tris pH 7.5, 5 % PEG4000
C5	0.1 M Tris pH 8.0, 20% PEG4000
D10	0.1 M Potassium chloride, 0.1 M HEPES pH 7.5, 15 % PEG6000
D11	0.1 M Tris pH 7.5, 15 % PEG6000
D12	0.1 M Tris pH 8.5, 20 % PEG6000
E8	0.1 M Magnesium acetate, 0.1 M MOPS pH 7.5, 12 % PEG8000
E9	0.2 M Sodium chloride, 0.1 M HEPES pH 7.5, 12 % PEG8000
F1	0.1 M HEPES pH 7.0, 20 % PEG8000
F4	0.1 M HEPES pH 7.0, 18 % PEG12000
G5	0.1 M Potassium chloride, 0.1 M HEPES pH 7.0, 1 M Ammonium sulfate

5.3.2. Structure determination of LpnE(73-375)

Of the various constructs, only LpnE(73-375) led to well-diffracting crystals. Its crystal structure was solved by molecular replacement and refined to a resolution of 1.75 Å. There are two molecules in the asymmetric unit. Each molecule consists of 17 antiparallel α-helices (Figure 5.8A). The helices are twisted into a right-handed super-helix. While short segments of the two molecules superimpose with root-mean-square deviation (rmsd) of 0.2–0.3 Å, the superposition of molecule A and B gives rmsd of 1.2 Å and indicates a small difference in super-helical twist of the helices (Figure 5.9). The loops connecting neighboring helices are

short on one edge of the super-helix and long on the other edge (Figure 5.8A). The helices are organized into helix-loop-helix pairs that are classified as Sel1-like repeats (SLRs). Two conventions have been used to describe multiple Sel1- like repeats. In a description by Lüthy et al. (Lüthy et al., 2004), the two helices of the Sel1-like repeat are joined by a short loop while the long loop connects two repeats. In the other description introduced by Mittl and Schneider-Brachert (Mittl and Schneider-Brachert, 2007), the two helices of the Sel1-like repeat are connected by a long loop. Here, we follow the first definition as it better fits the LpnE structure. The 17 α-helices of LpnE(73-375) form eight SLRs (α1–α16), with the C-terminal helix α17 capping the repeats (Figure 5.8D). The first helix of each SLR is designated as helix α_A and the second as helix α_B and are marked as n-α_A, n-α_B, where n = 1–8 and numbers the SLR in LpnE. The structural superposition of the eight repeats is shown in Figure 5.8B together with their aligned sequences (Figure 5.8C). The intra-SLR loops in LpnE are three residues long while the inter-SLR loops are longer, ~7 residues. SLRs in LpnE are 36 amino acids in length except for SLR1 which has 33 amino acids (Figure 5.8C).

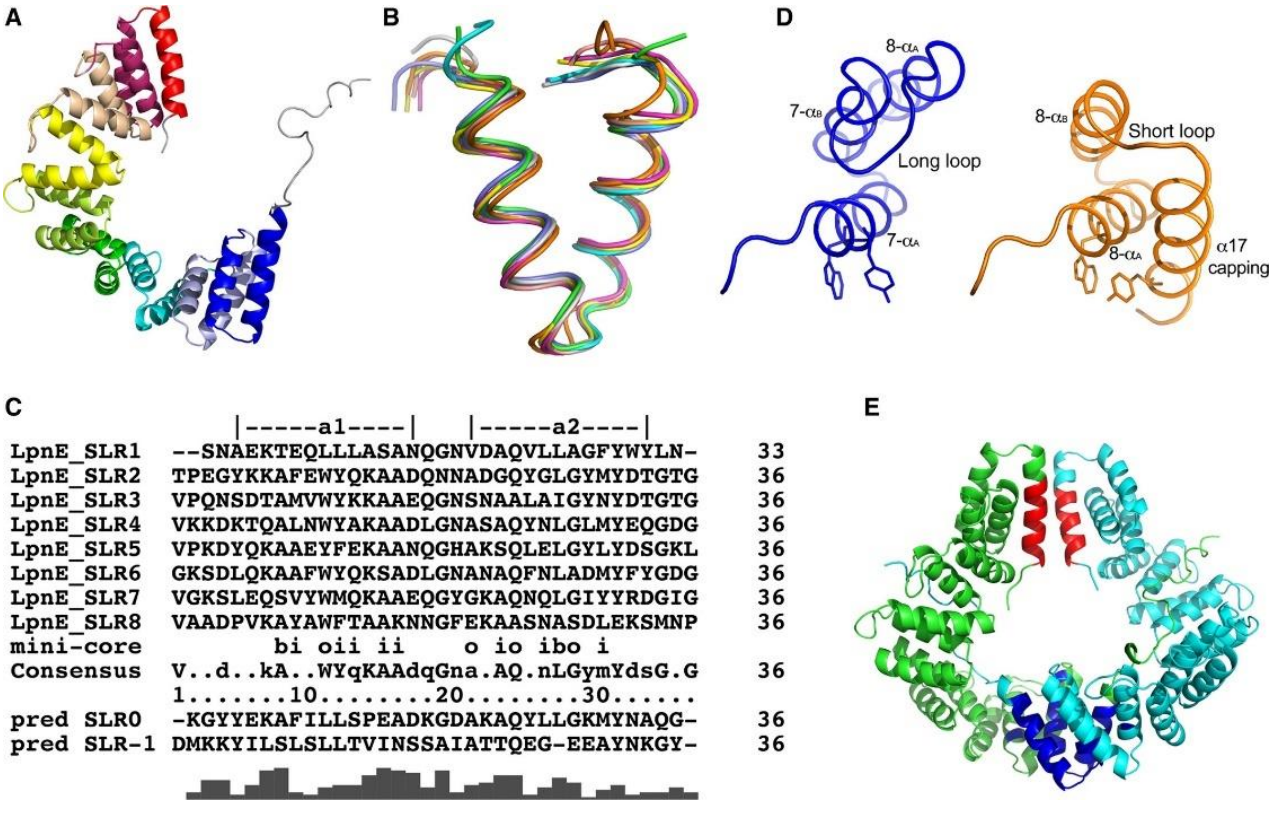

Figure 5. 8 Crystal structure of His-LpnE(73-375) comprising eight SLRs and a C-terminal capping helix. (A) Overall structure of His-LpnE(73-375). The SLR repeats are individually colored, with the N-terminal SLR1 blue and C-terminal SLR8 magenta. The capping helix is red. The N-terminal tag residues are colored gray; (B) structural superposition of SLR1 to SLR8 repeats of LpnE(73-375); (C) sequence alignment-based structural superposition of SLRs and the consensus sequence. The sequences of the predicted four helices in the 22–72 region are added to the alignment. The SLR0 sequence fits the consensus sequence quite well, but the sequence of putative SLR-1 diverges from the consensus; (D) the position of the capping helix differs from the expected position of the αA helix of the next SLR; (E) the packing of the two neighboring molecules involving the largest contact area. They interact through their N- (blue) and C-terminal (red) helices. The residues belonging to the affinity tag were omitted from this view.

The first helices (α_A) of Sel1-like repeats (SLRs) are located on the convex, exterior side of the super-helix, while the second helices (α_B) are on the concave, interior side. While both surfaces of the super-helix have hydrophilic character, they differ in that the convex side contains many charged side chains while the concave side is lined with uncharged polar side chains. Helix 1-α_A of the N-terminal repeat deviates the most from the SLR consensus sequence (Figure 5.8C). Hydrophilic residues Glu73, Lys74, Thr75, Glu76, Gln77, Ser82, and Asn84 are evenly dispersed throughout this helix. This arrangement comes in contrast with the amphipathic α_A helices found in the rest of the molecule, which display distinct hydrophobic and hydrophilic sides. The final C-terminal capping helix α17 does not follow the Sel1-like arrangement pattern. While the loop between 7- αB and 8- αA is seven residues long, the loop connecting 8-αB to α17 is only three residues long. Thus, α17, instead of adjoining helix 8-αB on the opposite side to 8-αA as occurs between neighboring SLRs, folds toward helix 8-αA and provides the side chain of Tyr367 for stacking with Trp329 that would otherwise be exposed to the solvent and caps the SLR repeats (Figure 5.8D).

Figure 5. 9 Superposition of the two independent LpnE molecules based only on SLR-1 and SLR-2.The difference in twist along the super-helix axis becomes clearly visible.

5.3.3. Crystal contacts

LpnE molecules in the asymmetric unit are related by non-crystallographic twofold symmetry and pack in a head-to-head/tail-to-tail fashion to form a twisted, doughnut-shaped dimer. Concave surfaces of each super-helix line the solvent-filled interior of the twisted doughnut (Figure 5.8E). The bulk of interactions between monomers are mediated by the N-terminal tag and connecting residues, which extend along the concave surface of the opposite chain. Indeed, these residues each contribute ~915 $Å^2$ of contact area to the ~2500 $Å^2$ interface between LpnE monomers. Elimination of the tag residues reduces the overall contact interface to ~655 $Å^2$, confirming the significance of this region to the interaction. Further inspection of contacts between untagged proteins reveals that ~364 $Å^2$ and ~279 $Å^2$ are contributed at the N- and C-terminus, respectively. Interactions between molecules in neighboring asymmetric units are even smaller than this, suggesting that untagged LpnE is monomeric. Consistent with

this notion, cleaved LpnE(73-375) runs only as a monomer on gel filtration, whereas tagged LpnE(73-375) shows both monomer and dimer species. The fact that non-protein residues give rise to dimerization suggests this oligomeric state is not of biological significance to LpnE.

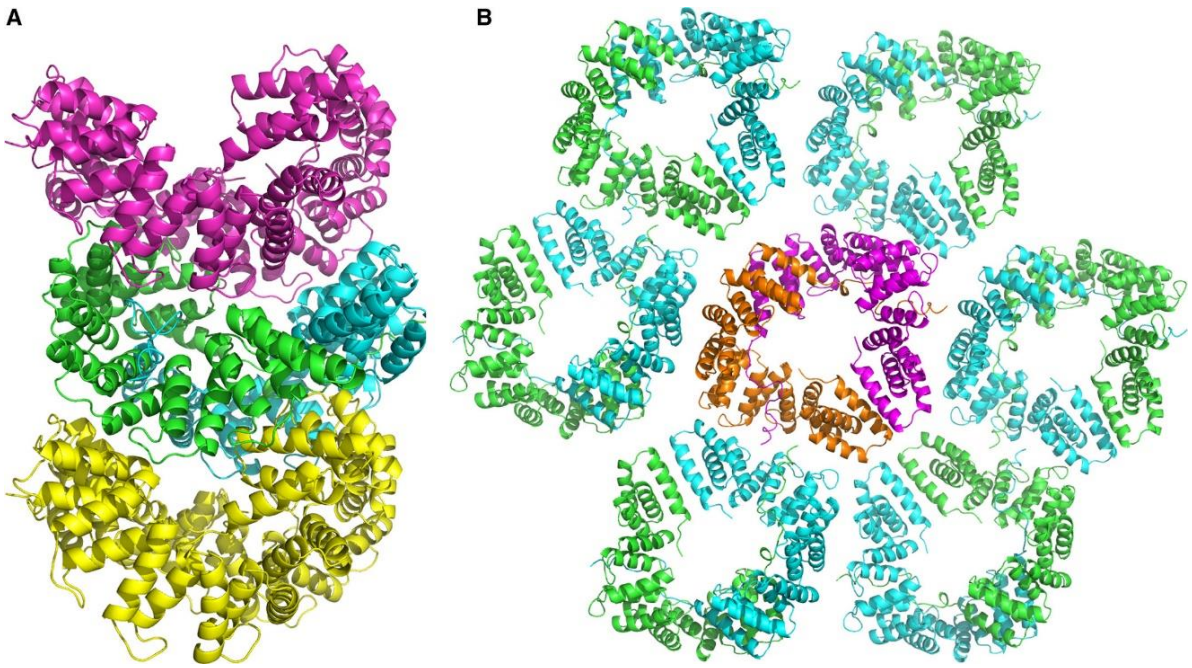

Figure 5. 10 Packing of LpnE(73-375) in the crystal. (A) The loose dimers form cylindrical stacks extending along the crystallographic a-axis with water channels in the middle; (B) the stacks are arranged in a hexagonal lattice in the b,c-plane viewed down the a-axis. This view is rotated 90° along the horizontal axis relative to the view in A.

The concave surface of Sel1-like repeat (SLR) superhelices conventionally mediates protein-protein interactions or substrate binding. As such, it is not altogether surprising that tag residues of a nearby chain nestle into this groove on LpnE. Residues that bind the hexa-histidine tag are located on helices 5-α_B, 6-α_A and 6-α_B. Three aspartate residues (Asp242, Asp275 and Asp281) stabilize the interaction by forming hydrogen bonds with sidechains of His50 and His54 at the N-terminus of the neighboring protein. Several tyrosine residues cushion the hexa-histidine tag by engaging in hydrophobic or hydrogen bond interactions. The remaining residues of the tag extend along a predominantly polar and uncharged surface formed by helices 2-α_B, 3-α_B, 4-α_A, 4-α_B and 5-α_B. Importantly, most connecting residues of the tag are poorly resolved in the electron density map. This conformational flexibility may arise from nonspecific interactions between polar tag residues and the concave surface of

LpnE. Surprisingly, Thr63, Glu64 and Asn 65 are better resolved than other residues of the extended chain. Closer inspection of this region shows that Glu64 is stabilized by hydrogen bonds to His230, Lys232 and Ser196, which anchor it to the concave surface of LpnE and rigidify flanking residues. Taken together, this interaction demonstrates the versatility of protein-protein interactions mediated by polar residues present on the concave surface of the LpnE super-helix.

LpnE dimers are stacked such that their interiors combine to form a solvent channel. The packing of these cylindrical stacks is pseudohexagonal, with seven molecules arranged into an approximate hexagon when viewed along the a-axis (Figure 5.10B). Histidine tags from separate stacks coordinate Ni^{2+} ions to stabilize the pseudohexagonal arrangement (Figure 5.11A/B). One of these Ni^{2+} ions is coordinated in a square planar fashion by ND1His52 and backbone amides of Met49, His50 and His51, with an average Ni-N distance of 1.95 Å (Figure 5.11C). The second Ni^{2+} ion connects NE2His52 and NE2His53 from neighboring chains, with water molecules above and below it to establish octahedral coordination (Figure 5.11D). These His-tag interactions provide a rationale for their role in crystal formation.

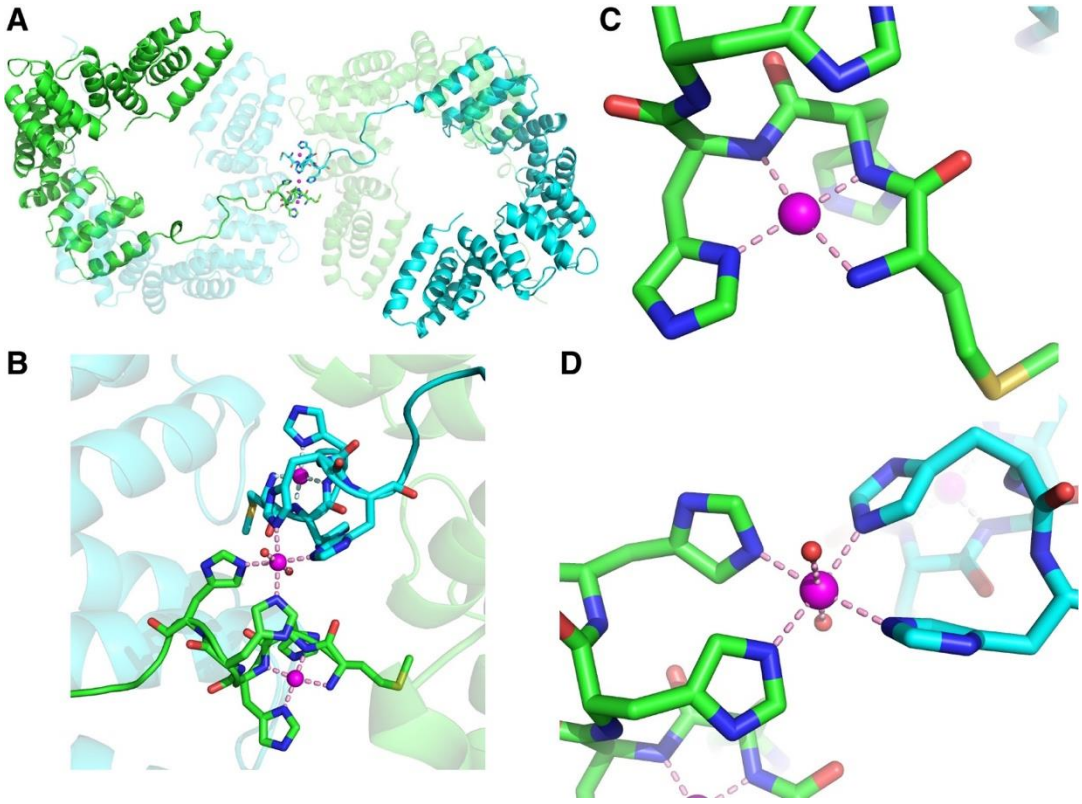

Figure 5. 11 The 'handshake' formed by the His-tags from two molecules. (A) The His-tags join two dimers shown in Figure B. The molecules contributing to the handshake are shown in wheat and cyan; their dimeric partners are semitransparent; (B) The close-up of the His-tags joined through tetrahedral coordination of a common Ni^{2+} ion. A second Ni^{2+} in each His-tag rigidifies the peptide through a square planar coordination with the His side chain and backbone atoms; (C) square planar coordination of Ni^{2+} ion within one His-tag; (D) tetrahedral coordination of the Ni^{2+} ion bridging two His-tags.

5.3.4. Cellular localization of LpnE

Previous intracellular studies of LpnE have focused on its role in vacuolar trafficking (Newton et al., 2007b) or host cell translocation (Bandyopadhyay et al., 2012), but not subcellular localization. To investigate this, we tagged LpnE(1-375) and LpnE(73-375) with a C-terminal GFP tag and overexpressed them in HEK293 cells. A subset of cells transfected with a vector encoding only GFP served as a control. The subcellular localizations of GFP and LpnE(1-375)-GFP in HEK293 cells were examined by fluorescence microscopy. The GFP control was distributed throughout the cells (Figure 5.12A, upper panels), whereas LpnE(1-375)-GFP predominantly showed a punctate pattern (Figure 5.12A, lower panels). The

subcellular localization of ectopically expressed LpnE(1-375) suggested that this protein may possess a hydrophobic membrane localization domain and bind-specific host organelles. To identify the nature of the punctate species, we analyzed the possible colocalization of ectopically expressed LpnE(1-375)-GFP with ER and Golgi markers in HEK293 cells. These experiments showed that LpnE(1-375)-GFP colocalizes with the *cis*-Golgi marker GM130 (Figure 5.12A, middle), but not the ER marker calnexin (Figure 5.12A, right) or 4′,6-diamidino-2-phenylindole (DAPI) identifying the nucleus (Figure 5.12A, left). To quantify the degree of colocalization of LpnE and intracellular marker, we calculated the Pearson correlation coefficient for each marker. A Pearson coefficient of 0.72 and 0.18 was obtained for GM130 and calnexin, respectively, supporting the colocalization of LpnE with Golgi. Conversely, overlap coefficients of 0.29 and 0.24 were obtained for GFP with GM130 and calnexin.

To identify the segment of LpnE that encodes for Golgi localization, we first investigated localization of the LpnE(73-375) fragment, which behaved well and was successfully crystallized. In contrast to the punctate pattern observed for LpnE(1-375), LpnE(73-375) was distributed throughout the cell (Figure 5.12C), indicating that the first 72 residues are guiding Golgi localization. Next, we investigated the localization of a predicted mature protein: LpnE(22-375). Surprisingly, this construct was also distributed throughout the cytosol of the HEK293 cells (Figure 5.12D). Thus, neither of the ectopically expressed LpnE(73-375) or LpnE(22-375) colocalized with the *cis*-Golgi marker. In order to test whether the first 22 residues of LpnE serve as a localization peptide, we fused this LpnE peptide to the N-terminus of two other effector proteins LegA15 and LpiR1, which we previously showed to have cytosolic localization. We expressed these proteins in HEK293 cells and observed their localization. Wild-type LegA15 and LpiR1 were distributed throughout the cytoplasm. When these proteins were fused to the LpnE peptide, however, they displayed a punctate pattern that colocalized with the Golgi marker (Figure 5.12E). The Pearson correlation coefficients for wild-type LegA15 and Lpir1 with *cis*-Golgi marker GM130 were 0.34 and 0.29, respectively, while for the tagged proteins, it was 0.71 and 0.53, respectively (Figure 5.12F). These results suggest that the putative signal peptide of LpnE is responsible for its *cis*-Golgi localization.

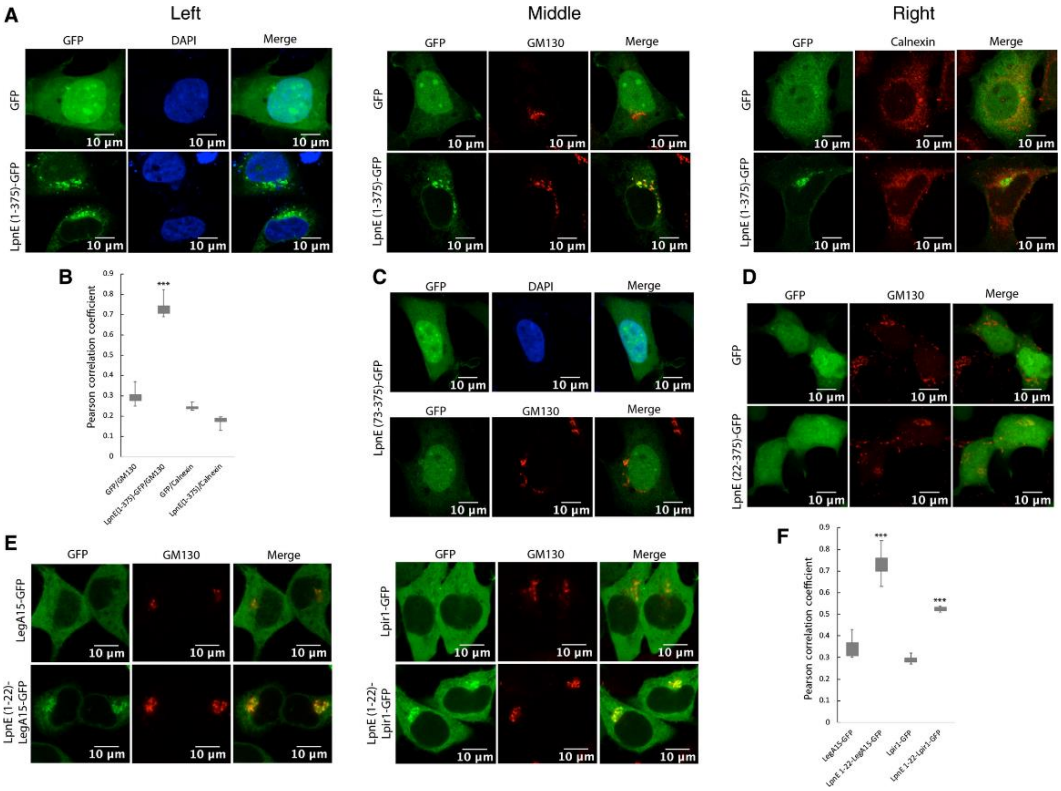

Figure 5. 12 Localization of various LpnE fragments in HEK293 cells. (A) The plasmid containing either GFP or LpnE(1-375)-GFP was transfected to HEK293 cells. Left grid—nucleus visualized with DAPI. Middle grid—the *cis*-Golgi visualized with anti-GM130 antibody, right grid—the ER visualized with anti-calnexin antibody. Top panels—GFP alone, lower panels—LpnE(1-375)-GFP. The scale bar indicates the length of 5 µm; (B) quantitation of colocalization by Pearson coefficient. The Pearson correlation coefficient was determined with coloc 2 plugin of the imagej software 35 and was calculated with at least 30 cells (mean and standard deviation, SD, of the mean from each experiment, ***P < 0.001); (C) LpnE(73-375)-GFP localization in HEK293 cells compared with DAPI staining of the nucleus (top panel) and the cis-Golgi staining with anti-GM130 antibody (lower panel); (D) LpnE(22-375)-GFP localization in HEK293 cells compared with GM130 marker; (E) fusion with LpnE signal peptide led Legionella pneumophila effector LegA15 and Lpir1 to Golgi apparatus. (F) quantitation of colocalization by Pearson coefficient of Figure 5.12D with at least 30 cells was calculated (mean and standard deviation, SD, of the mean from each experiment, ***P < 0.001).

5.4. Discussion

LpnE is one of several *Legionella* effectors containing predicted SLR-like repeats and shares ~30% sequence identity with EnhC, LidL, Lpg1062, and Lpg1356. The effector plays a role in bacterial entry into the host cell, in phagosome acidification and helps the LCV evade fusion with the lysosome. Two host proteins, OCRL and OBSL1, were confirmed to be LpnE-

interacting partners and the LpnE-binding regions on these two proteins have been roughly identified (Newton et al., 2007b; Weber et al., 2009). We determined which segment of LpnE is directing its cellular localization, determined the 3-D structure of the SLR repeat containing domain, and refined the OCRL segment required for interaction with LpnE.

5.4.1. LpnE localization

Our localization studies showed that full-length LpnE is localized to the *cis*-Golgi in HEK293 cells. We then investigated if the SLR-containing fragment, LpnE(73-375), retains this localization pattern. Intriguingly, this construct was found predominantly in the cytosol. Since the first 22 residues were predicted as a signal peptide, we rationalized that the localization is associated with the segment 22–72. This is not the case, however, as LpnE(22-375) was also found in the cytosol. We, therefore, concluded that the first 21 amino acids direct LpnE to the *cis*-Golgi. LpnE is not delivered to the host cell through the Dot/Icm secretion system and the mechanism of its translocation is presently unknown (Newton et al., 2007b). The N-terminus does indeed contain a signal peptide, but it is not known if this peptide is recognized inside *Legionella* or in the host cell and if or when it is cleaved off. While the mechanism of LpnE retention at the *cis*-Golgi remains unknown, it is possible that specific amino acid residues act as a sorting signal. Specific recruitment of LpnE to the *cis*-Golgi may depend on a preference for characteristic glycerophospholipid/sphingolipid ratios, as these lipids distinguish *cis*-Golgi from *trans*-Golgi (Banfield, 2011).

The biological significance of LpnE(1-375) retention at the *cis*-Golgi remains unclear. It is important to note that a punctate localization pattern was observed only in ~20% of HEK293 cells, while a cytosolic localization was observed in the remaining cells. This suggests cleavage of the signal peptide in some instances. LpnE is known to bind phosphatidylinositol-3-phosphate (PI3P) and localize to the *Legionella*-containing vacuole (LCV) (Weber et al., 2009). The retention of LpnE at the LCV may be mediated by the signal peptide and/or the aa 23–72 segment. Cleavage of the signal peptide would liberate LpnE from the LCV, allowing it to interact with various eukaryotic proteins via the concave surface of its super-helix. Recruitment of OCRL to the LCV and its association with LpnE may render the latter

inaccessible to host signal peptidases and thereby impede the infective role of LpnE. This notion is in keeping with previous studies showing that the first 51 residues (including the signal peptide) are not necessary for LpnE export and invasion (Newton et al., 2007b). In short, LpnE may interfere with host cell processes outside of its known role in uptake, with the signal peptide allowing its brief retention on the LCV.

5.4.2. LpnE/OCRL interaction studies

Our SEC-MALS data suggest that both LpnE(73-375) and LpnE(22-375) interact with OCRL(10-208), albeit with different stoichiometry. The 69.6 kDa peak containing both LpnE(73-375) and OCRL(10-208) agrees reasonably well with the theoretical molecular weight estimate of 56.6 kDa for a 1 : 1 complex of these proteins. Meanwhile, the 147.7 kDa peak observed for LpnE(22-375)-OCRL(10-208) cannot be rationalized in the same way, as a 1 : 1 stoichiometry would correspond to a theoretical molecular weight of 62.3 kDa. Thus, the observed peak at 147.7 kDa may represent an LpnE(22-375)$_2$-OCRL(10-208)$_2$ heterotetramer.

It is tempting to speculate from these data that the N-terminal 22–72 residues of LpnE promote its own dimerization and thereby render a dimer-of-dimers in the presence of OCRL(10-208). This hypothesis, however, would also suggest that LpnE(22-375) readily dimerizes in the absence of OCRL(10-208). Our studies of LpnE(22-375) have shown this not to be the case, as LpnE(22-375) appears to be monomeric by SEC-MALS (39.4 kDa calculated vs. 41.8 kDa measured). An alternative explanation is that the binding of OCRL(10-208) to LpnE(22-375) induces a conformational change in either LpnE or OCRL, exposing a protein–protein interaction interface that gives rise to dimerization. The fact that no dimer-of-dimers is observed for LpnE(73-375)-OCRL(10-208) may point to a special role for the N-terminal 22–72 residues of LpnE in promoting this process.

LpnE(22-375) produces a broad peak when eluting from Biorad SEC650 or GE Superdex200 columns. This indicates that LpnE(22-375) is conformationally heterogeneous at the concentrations used for this assay. Indeed, a 100 µL injection of LpnE(22-375) at only 2 mg/ml was used for the SEC-MALS analytical column, whereas 500 µL injections at concentrations above 15 mg/ml were the norm during routine purifications. Since LpnE(73-375) retains distinct monomer and dimer peaks at high concentrations, the conformational

heterogeneity of LpnE(22-375) must be attributed to its N-terminal 22-72 residues. We expect some level of flexibility for this fragment of LpnE, as it was accessible to trypsin. Still, flexibility alone does not explain the concentration-dependent change in gel filtration profiles for LpnE(22-375). It is possible that the N-terminal 22-72 residues interact transiently with the concave surface of the SLR super-helix in a nearby molecule. Such an interaction would lead to the formation of higher order oligomers that rapidly assemble and disassemble when the protein concentration is high enough. This explanation is consistent with previous findings that show TRP repeat proteins to be promiscuous in their binding properties.

Copurifying LpnE(22-375) with OCRL(10-208) produces a sharp peak on gel filtration with purified fractions showing low polydispersity on DLS. This suggests that OCRL(10-208) limits the conformational diversity of LpnE(22-375). A potential explanation for this phenomenon is that OCRL(10-208) occupies the protein-binding interface of LpnE(22-375) and prevents the formation of LpnE oligomers at high concentration. OCRL(10-208) may also stabilize the LpnE N-terminus by interacting with residues in this region. These effects could rationalize the crystallization of LpnE(22-375) when purified in the presence of OCRL(10-208). That is, OCRL(10-208) may retain LpnE(22-375) in a single conformation that is amenable to crystallization. It is worth noting, however, that OCRL was absent from crystals obtained using preparations of LpnE(22-375)-OCRL(10-208). The gradual disappearance of OCRL(10-208) from these preparations may facilitate the crystallization of LpnE(22-375) by increasing the concentration of properly folded LpnE monomers in solution. Alternatively, a small fragment of OCRL(10-208) may remain bound to LpnE(22-375) to stabilize it for crystallization. Since no structure was solved for LpnE(22-375), it remains unclear what role OCRL(10-208) played in the process of crystallization. Still, the fact that LpnE(22-375) crystals were obtained only when this protein was copurified with OCRL(10-208) supports the notion that OCRL stabilizes LpnE by a direct interaction.

His-LpnE(22-375) crystallized in a variety of wells from the ProComplex Suite (Qiagen) (Table 1), which has conditions derived from those used in the protein data bank (PDB) for protein-protein complexes. The screen is biased in favor of low to medium molecular weight PEGs, 0.1 – 0.2 M salt concentration and pH ranges from 5.0 – 8.0 (with a few exceptions). Inspection of Table 1 shows that His-LpnE(22-375) prefers a pH range of 7.0 – 8.5 and medium

molecular weight PEGs (4, 6, 8 and 12K) at 15 – 20 %. Conditions C5, D11, D12, F1 and F4 contain no salt, indicating that this component is dispensable for the crystallization of His-LpnE(22-375). It is intriguing that crystals of His-LpnE(73-375) can be grown only in conditions previously used to obtain protein-protein complexes, because LpnE(22-375) was purified in the presence of OCRL(10-208). That is, we were hoping to obtain a complex of LpnE(22-375) and OCRL(10-208) when crystals of His-LpnE(22-375) were grown. Protein-protein complexes are sensitive to the destabilizing effect of extreme pH. Furthermore, complexes are often less soluble than their individual constituents and therefore require lower molecular weight PEGs and PEG concentrations (Radaev et al., 2006). His-LpnE(22-375) crystallizes at moderate pH values and with lower % PEG than His-LpnE(73-375) [0.1 M MES pH 6.5, 0.2 M ammonium sulfate, 30 % PEG 8000]. These observations, when coupled to the fact that His-LpnE(22-375) could not be crystallized in the absence of OCRL(10-208), can be used to further justify our suspicion that some component of OCRL interacts with and stabilizes LpnE to promote crystallization. It is difficult to speculate what role the His-tags play in the crystallization of LpnE(22-375), although they may promote the interaction of nearby molecules in a similar fashion to that observed in our model of LpnE(73-375). Future efforts to crystallize LpnE(22-375) should continue to explore the stabilizing effect of OCRL(10-208) on this protein.

To investigate which parts of LpnE and OCRL are essential for their interactions, we expressed a series of LpnE constructs with N-terminal deletions and OCRL constructs with C-terminal deletions. We showed that while OCRL(10-208) and OCRL(10-183) bind LpnE, the slightly shorter construct lacking the polyproline sequence, OCRL(10-176) binds LpnE very weakly or not at all. This suggests that the polyproline segment of OCRL is critical for LpnE binding. The polyproline motif within OCRL likely adopts a left-handed PPII helix, as this is energetically favorable to PPI, and these helices are implicated in protein–protein interactions (Moradi et al., 2009). This extended structure could fit alongside the Sel1 helical motifs forming a concave surface on LpnE. To determine the region of LpnE critical for binding OCRL, we investigated binding of OCRL(10-208) to the shortened LpnE constructs. LpnE(158-375) is missing five helices from the crystallized LpnE construct and still binds OCRL. Moreover, like LpnE(73-375), this construct also binds OCRL(10-183) but not the

shorter OCRL constructs. However, deleting one more helix from LpnE [LpnE(176-375)] abrogates binding, indicating that the 159–175 helix (3-α_B) is critical for the interaction. This helix is at the center of the curved surface in the structure of LpnE. Taken together, our data suggest that the interaction is between the polyproline segment of OCRL and the 3-α_B helix of LpnE. Our studies make clear that the N-terminal PH domain of OCRL is unlikely to be involved in binding LpnE and the primary interactions are predominantly with the unstructured linker between the PH and 5-phosphatase domains of OCRL.

The polyproline motif in OCRL comprises residues 177 – 183 and consists of PPPPPFS. Since no crystal structure was obtained for LpnE in complex with OCRL, it remains unclear how the polyproline motif interacts with the 3-α_B helix of LpnE. Our structure of LpnE(73-375) provides an example of how an extended peptide can bind the concave surface of the super-helix. Inspection of the LpnE(73-375) crystal structure reveals that tag residues of one chain are in proximity to the 3-α_B helix of a neighboring chain. The PISA server shows that interacting residues in this helix consist of Asn160, Leu163, Ala164, Tyr167, Asp170 and Thr171. These residues have sidechains oriented toward the center of the doughnut shaped dimer and characterize the concave surface at this region. Polyproline motifs engage in protein-protein interactions because they lack intramolecular hydrogen bonds and rely on external sources to satisfy them. Considering that several of the interacting residues in 3-α_B have polar sidechains that can act as hydrogen bond donors, it is easy to envision how this helix might stabilize a polyproline II helix. It should also be noted that the polyproline motif is highly hydrophobic. Thus, the presence of hydrogen bond donors may need to be counterbalanced with nonpolar residues that contribute to hydrophobic packing of the proline sidechains against the super-helix. Leu163, Ala164 and the aromatic ring of Tyr167 could serve this purpose, although these residues occupy only a small central segment of the 3-α_B helix and are unlikely to interact with five consecutive proline residues in an extended conformation. It should not be assumed, however, that the polyproline motif binds exclusively to residues within the 3-α_B helix. Indeed, Ala182, Tyr186, Ala189 and Ala190 along the concave face of the neighboring helix, 4-α_A, might form a hydrophobic cleft with the 3-α_B helix. If residues from both these helices work in concert to facilitate the binding of OCRL, we should expect that removal of 3-α_B would abolish the interaction. This is consistent with our observation that binding of

LpnE(176-375) to OCRL(10-208) does not occur. A complex structure of LpnE with the OCRL polyproline peptide will still be needed to truly appreciate the mode of interaction between these proteins.

5.4.3. Structural comparison with other Sel1-like repeat proteins

SLR proteins are prevalent in nature and many structures of proteins containing these domains are known. Comparison of LpnE(73-375) with other such proteins using the DALI server (Holm and Rosenström, 2010) identified EsiB (PDB id 4BWR (Urosev et al., 2013)) from extra-intestinal pathogenic *E. coli* (ExPEC), HcpB (PDB id 1KLX (Lüthy et al., 2002)), HcpC (PDB id 1OUV (Lüthy et al., 2004)) of the *Helicobacter pylori* cysteine-rich (Hcp) family and mouse SEL1L (PDB id 5B26 (Jeong et al., 2016)) of the ER-associated protein degradation (ERAD) machinery as the closest structural homologs. EsiB, HcpB, and HcpC are all SLR proteins from bacterial pathogens and their roles in virulence remain uncertain. Although LpnE has greater sequence identity with EsiB and SEL1L (36%) than with HcpB (29%) or HcpC (31%), structural superposition shows that repeats align best with HcpC (RMSD 2.3 Å). This difference is attributable to the long inter-repeat loops, which lean in the direction of the convex outer face of the super-helix. These longer loops play an important role in defining the super-helical geometry of the structure and are noticeably better aligned between LpnE and HcpC. It is likely that this structural similarity made HcpC a good candidate model to obtain initial phases for LpnE by molecular replacement.

Like EsiB, LpnE lacks the intra-repeat disulfide bonds seen in HcpC or its shorter homolog, HcpB to help stabilizing the repeats. Instead, the super-helical packing of LpnE is stabilized predominantly by its continuous hydrophobic core. This stabilization strategy is also observed in EsiB, which exhibits a strikingly similar pattern of SLR residues to that seen in LpnE. Despite sharing the same sequence identity with LpnE as EsiB, SEL1L forms a super-helix with a tighter twist and associates into dimers through C-terminal helical extension following the last SLR repeat.

In EsiB, the hydrophobic core is centered at Leu7 on helix A. Leu7 makes contacts with Trp28 and hydrophobic residues found at positions 26, 29, 30 and 33 of helix B. In the same way, Leu7 of LpnE is nearly always in close contact with Trp27, and hydrophobic residues

found at positions 25, 28, 29 and 32. The presence of Trp27 in LpnE as an analogue to Trp28 of EsiB is nontrivial; this residue has been implicated in substituting for the effect of disulfide bonds in Hcp family SLR proteins. Additional similarities between the hydrophobic cores of LpnE and EsiB include contacts between Leu7 and Tyr11, and hydrophobic interactions between positions 10 and 25 in LpnE (10 and 26 in EsiB).

A notable difference between LpnE and EsiB can be seen at position 29. In LpnE, this position is dominated by the residues glutamine and glutamate, whereas the corresponding position 30 in EsiB is predominantly arginine. The hydrogen bond between guanidinium and hydroxyl groups of Arg30 and Tyr11 in EsiB is not mimicked by Gln/Glu29 of LpnE. Thus, the kink observed in helix B of EsiB is not present in LpnE.

Intra-repeat interactions are stabilized by hydrogen-bonded glutamine residues at positions 4 and 34 (35 in EsiB) in both LpnE and EsiB. The stacking interaction between Tyr11 and Gln24 of EsiB is seen in only two of the eight repeats of LpnE. Position 23 of LpnE is dominated by lysine, rather than glutamine. The intra-repeat loops are comprised mostly of seven residues, with SLR2 having one residue less, and SLR5 lacking two residues. As with EsiB, a critical valine (Val17 in LpnE) mediates interactions with the hydrophobic core residues Tyr11 and 26. Interestingly, this valine is present in only four of seven intra-repeat loops observed in our LpnE structure. The inter-repeat loop is comprised of three residues in LpnE and EsiB.

Secondary structure predictions for LpnE indicate the presence of four helices preceding the domain for which we have determined the structure (aa 1-72). Of these, helices α3 and α4 conform well to the consensus sequence of the SLR repeats in LpnE (Figure 5.8D) and we suggest that they indeed form an additional SLR repeat (predicted SLR0). Trypsin cleavage of LpnE occurs within a nine-residue linker connecting SLR0 to SLR1 (the N-terminal repeat observed in our structure). The sequences of the predicted helices α1 and α2 do not correspond well to the consensus, and moreover, helix 1 falls within the predicted signal peptide. We, therefore, conclude that only helices 3 and 4 form a SLR repeat and that there are consequently nine SLRs in LpnE.

5.4.4. Ni^{2+} coordination by the LpnE hexa-histidine tag

The presence of the His-tag promoted crystallization of LpnE. The six histidines of the tag acquired 1.5 Ni^{2+} ions per tag. One Ni^{2+} ion rigidifies the conformation of the tag while the other is shared between the tags from two different molecules and helped in crystal packing. These two Ni^{2+} adopt two different coordination environments, the intratag square planar coordination with Ni-N distances of 1.95 Å and the intertag octahedral coordination with Ni-N/O distances of 2.13 Å. The observation of two different coordination types in one protein crystal is rather unusual, but it appears to be serendipitous and unrelated to LpnE itself.

Table 5. 2 Data collection and refinement statistics for His-LpnE(73-375)

Data collection	
Space group	$P2_12_12_1$
Unit cell (Å)	67.7, 84.1, 128.3
Resolution (Å)	48.77–1.75 (1.81–1.75)
Total reflections	1 058 671 (135 933)
Unique reflections	74 365 (11 788)
Completeness (%)	99.83 (98.68)
Redundancy	14.2 (11.5)
Mean I/σ(I)	21.12 (1.85)
R_{merge} (%)	8.3 (148.9)
CC1/2	100 (73.4)
Refinement	
Resolution (Å)	48.77–1.8 (1.86–1.80)
No. of reflections	68 575 (6782)
R_{work} / R_{free} (%)	18.7 / 21.4
No. of atoms / waters	5361 / 307
RMSD from ideal values	
Bond lengths (Å)	0.013
Bond angles (°)	1.39
Ramachandran plot (%)	
Favored	97.37
Allowed	1.70
Outlier	0.93
PDB ID	6DEH

6. Crystal Structure of the *Legionella* Effector MavE

6.1. Literature Review on MavE

MavE (*lpg2344*) is one of many putative *Legionella pneumophila* effectors recognized as a substrate of the Icm/Dot T4SS. In the study under which MavE was identified, a C-terminal signal sequence was found in effectors restoring LCV colocalization to a translocation deficient SidCΔ100 (lacking its own C-terminal 100 residues). Effectors identified in this way were termed either Rav (Regions allowing vacuole colocalization) or Mav (More regions allowing vacuole colocalization) (Huang et al., 2011).

MavE contains a C-terminal transmembrane helix (aa 187 – 203) and lacks the E-block motif seen in ~50% of known translocated substrates (Huang et al., 2011). Our collaborators have shown that MavE localizes to a micro-domain on the cytosolic side of the LCV (unpublished data). This localization pattern can likely be attributed to the C-terminal transmembrane helix in MavE. Secondary structure predictions show MavE to be fully α-helical with a short N-terminal helix (aa 2 – 11) flanked by an unstructured region extending to Ser28. Analysis of the MavE protein sequence by PsiBLAST (Position-specific iterated Basic Local Alignment Search Tool) (Altschul et al., 1997) shows many homologues with start sites corresponding to Met38. This may represent a mis-annotation of the MavE start site listed in UniProt, as homologues belonging to different bacteria were found to align at Met38, while only hypothetical proteins from L. pneumophila aligned to Met1 of MavE.

A direct interaction between MavE and the *Legionella* SNARE (Soluble NSF(N-ethylamide sensitive factor) Attachment protein REceptor) mimic (Campodonico et al., 2016; Shi et al., 2016) and LegC7, has been identified by yeast two-hybrid assay and the LUMIER assay. The latter approach allows the detection of protein-protein interactions in mammalian cells. Briefly, FLAG-tagged bait and *Renilla* luciferase-tagged prey proteins were co-transfected into HEK293 cells. Lysates were added to anti-FLAG plates and co-precipitation was evidenced by luminescence. MavE was also found to suppress the growth defect observed in *Saccharomyces cerevisiae* overexpressing LegC7. These findings led MavE to be described as an effector of effectors (or meta-effector) (Urbanus et al., 2016). Meta-effectors promote

the regulatory complexity that is essential to the *Legionella* life cycle, although it remains unclear why MavE would inhibit LegC7.

6.2. Experimental Procedures for the MavE Project

6.2.1. Cloning of recombinant MavE

The MavE (*lpg2344*) gene was amplified from *Legionella pneumophila* (Philadelphia) genomic DNA by PCR. Residues 183 – 204 are predicted to comprise a transmembrane (TM) region (Program Phobius). To clone only the soluble domain of MavE, we amplified the DNA sequence encoding residues 2 – 172. This construct terminates just after the final hydrophilic helical stretch and excludes the following loop and TM region. The MavE(2-172) insert DNA sequence was placed into pMCSG7 and pRL652 vectors by ligation independent cloning (LIC), incorporating an N-terminal TEV-cleavable His_6- or GST-tag, respectively. His_6-MavE(2-172) expressed poorly in BL21(DE3)pLysS and GST-MavE(2-172) did not readily bind the glutathione resin. Running PsiBLAST on MavE showed that most homologous proteins have start sites corresponding to residue M38. To explore the possibility of a misannotated start site, we amplified MavE(39-172) from the plasmid containing MavE(2-172). As previously described, the MavE(39-172) DNA insert was placed into pMCSG7 and pRL652 vectors via LIC using the extensions underlined above. This construct of MavE was used for expression, purification and crystallization trials.

6.2.2. Protein expression and purification

The expression of His_6-MavE(39-172) was carried out as described in section 4.2.3 for LpnE with the following exceptions: Once purified on Qiagen NiNTA beads (Figure 6.1B), His_6-MavE(39-172) was concentrated to 16mg/ml and digested overnight at room temperature with TEV protease at a 1:50 enzyme/substrate molar ratio. The extent of digestion was monitored by SDS-PAGE (see section 4.5.1). Untagged MavE was then loaded onto a Biorad SEC70 or GE SEC75 column for buffer exchange and further purification. MavE eluted from

the size exclusion column in a single peak corresponding to monomeric protein. Peak fractions were collected and concentrated to 25 mg/ml for crystallization (Figure 6.1A/C).

A seleno-methionine derivative of MavE(39-172) was produced by inhibiting methionine biosynthesis immediately prior to induction. Specifically, 100 mg lysine, phenylalanine and threonine and 50 mg isoleucine, leucine and valine were added to 1 L of culture 15 minutes prior to induction. 60 mg L-seleno-methionine was also added to allow the incorporation of selenium atoms into overexpressed MavE during induction.

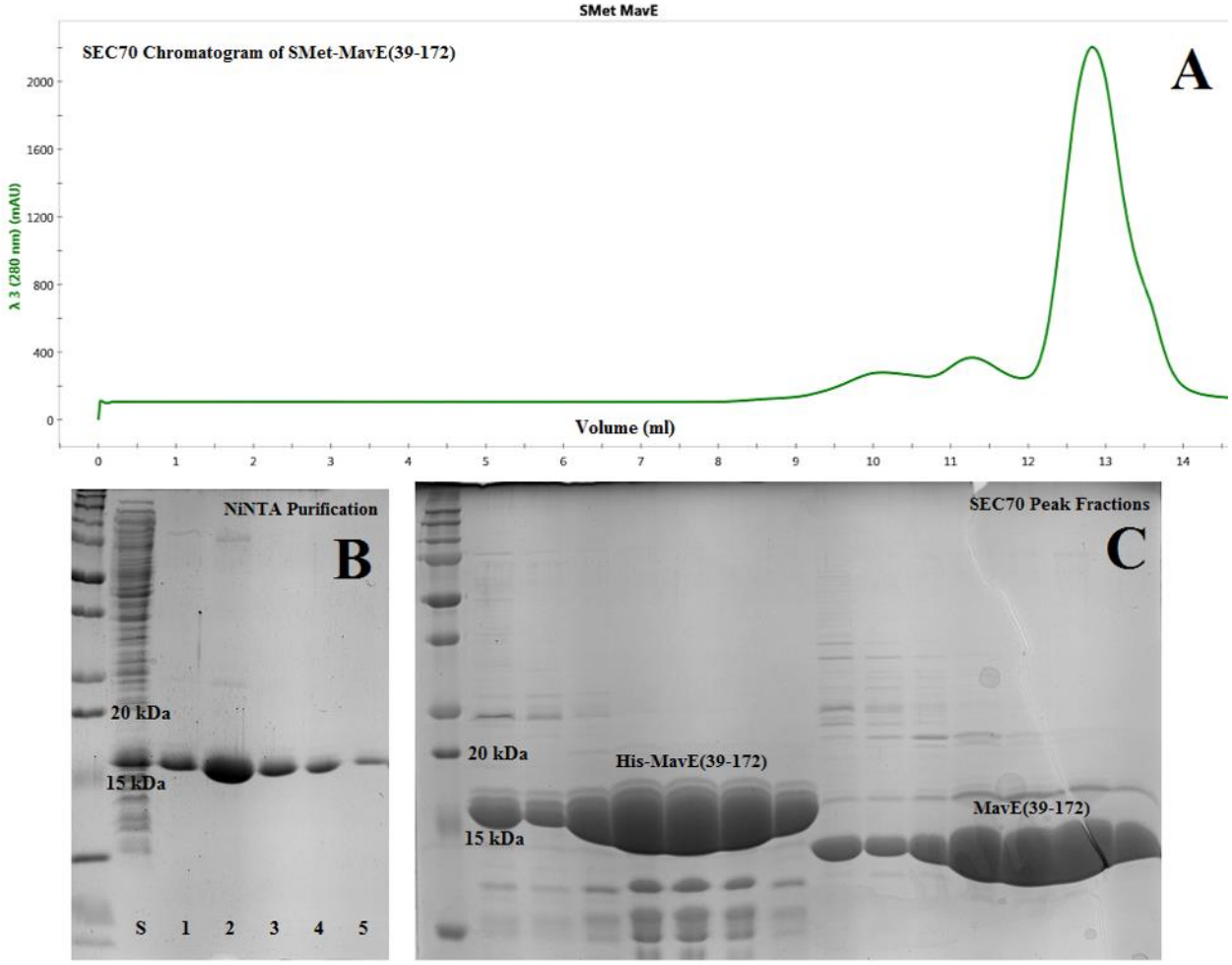

Figure 6. 1 Purification of His-MavE(39-172). (A) Size exclusion chromatogram of seleno-methionine MavE(39-172) purified on a Biorad SEC70 column. (B) Purification of His-MavE on NiNTA. S = soluble protein (clarified lysate), 1 – 5 = elutions 1 – 5. Protein is purified in 15 mM Tris, pH 8.0, 50 mM NaCl and 100 mM imidazole. (C) Peak fractions collected from size exclusion on His-MavE(39-172) and TEV-cleaved MavE(39-172) (depicted in A).

6.2.3. Crystallization of MavE(39-172)

Both His_6-MavE(39-172) and MavE(39-172) were screened for crystallization using Crystal Screen HT, Index (Hampton Research, Aliso Viejo, CA), JCSG Core II and Classics Suite (Qiagen, Toronto, Canada). His_6-MavE(39-172) did not crystallize under any of the conditions tested, whereas MavE(39-172) produced crystals under several conditions. After optimization by the hanging drop vapor diffusion method, the best crystals were obtained at 20 °C in drops containing 1 µl protein in 15 mM Tris-HCl, pH 8.0, 50 mM NaCl mixed with 1 µl of reservoir solution (10% PEG 20000, 0.1 M Citrate, pH 3.0) and suspended over 500 µl reservoir solution.

6.2.4. Data collection and structure determination

The protein crystals were cryo-protected by transferring to 1 µl mother liquor containing 20% (v/v) ethylene glycol. Diffraction data were collected to 1.8 Å at the Canadian Macromolecular Crystallography Facility (CMCF) 08ID beamline, Canadian Light Source, using a MAR300CCD Detector (Grochulski et al., 2011). Integration and scaling were carried out using the XDS software package (Kabsch, 2010) (autoprocess). MavE(39-172) contains only one methionine at position 51 and substitution of this residue for seleno-methionine produced enough anomalous signal to solve the structure by single anomalous dispersion (SAD) using the phenix.autosolve script. Refinement of the structure was carried out using phenix.refine (Adams et al., 2011).

6.3. Results

6.3.1. Overall structure

The crystal structure of MavE(39-172) was solved by single anomalous dispersion (SAD) and refined to a resolution of 1.8 Å. MavE is entirely α-helical with a right-handed up-down-down-up-down topology. These five helices, labelled A, B, C, D and E, are composed of residues 39 – 58, 61 – 70, 94 – 114, 120 – 141 and 150 – 167, respectively (Figure 6.1). The

three molecules in the asymmetric unit are designated *A*, *B* and *C*. Molecules *B* and *C* are related by twofold symmetry, and molecule *A* is also related by a twofold axis to an *A* molecule in a neighboring unit cell (Figure 6.2). This results in a crystal lattice comprising layers of *BC* and *AA* dimers. Molecules within the unit cell are arranged such that two *BC* layers are sandwiched between a single layer of *A* molecules (Figure 6.2). The N- and C-terminal residues of MavE are near one another, with Arg155 and Arg162 in helix D forming hydrogen bonds to Leu57 and Glu54 in helix A, respectively. Helices B and C are connected by a 22-residue loop, with the first 10 residues having poorly defined electron density. This loop stretches across a stable core of the protein formed by helices A, C, D and E.

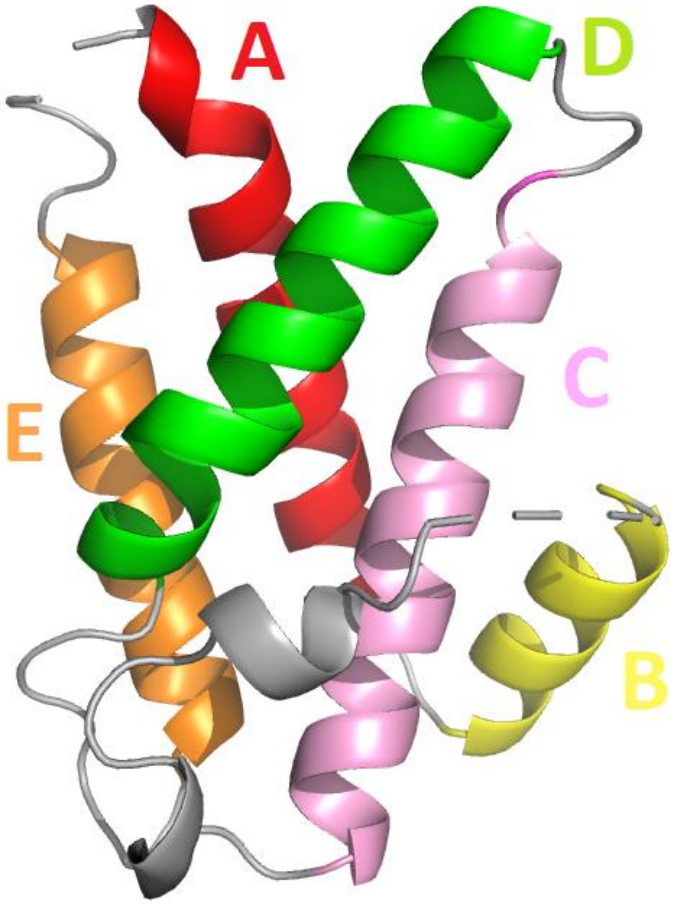

Figure 6. 2 Overall structure of MavE(39-172).Helices A, X, B, C, D are coloured in red, yellow, green, orange and pink (respectively). Loop regions are shown in grey.

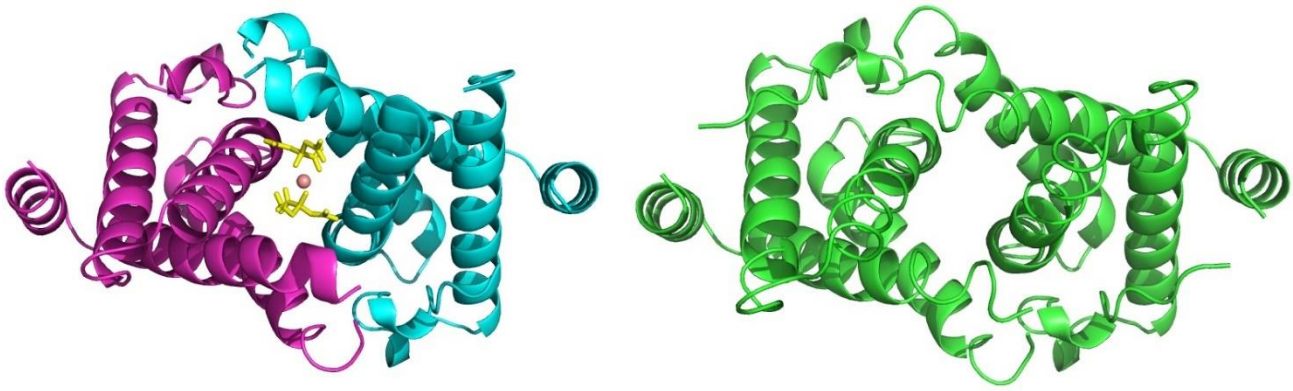

Figure 6. 3 *BC* and *AA'* interfaces in the MavE(39-172) structure. Chains A, B and C are shown in green, cyan and magenta, respectively. The dimer arrangement is the same for both *AA'* and *BC* dimers, although two citrate molecules coordinate a sodium ion in the latter. A small molecule of unknown identity is present at the same region of the *AA* interface, as evidenced by the F_o - F_c map. Interacting proteins are not shown as organized in the crystal lattice.

6.3.2. Crystal Contacts

Significant intermolecular contacts in the MavE crystal occur between *B* and *C* chains or *A* and *A'* from a neighboring unit cell. Each of these ~1000 $Å^2$ interfaces (https://www.ebi.ac.uk/pdbe/pisa/) is formed by N- and C-terminal residues, which contact a kinked region of helix C in the symmetry-related molecule. Two citrate molecules at the center of the *BC* interface stabilize their association through a network of water molecules. Residues contributing to this interaction are Ala38, Glu42, Gln125 and Ser128 of each chain. The positive difference map ($mF_o – DF_c$) indicates the presence of another stabilizing element at the *AA'* interface, although the identity of this molecule remains unclear.

The *BC* and *AA'* connections are likely an artifact of crystallization, as MavE is monomeric in solution. Furthermore, each interface relies on a stabilizing element within a solvent filled cavity. In *BC*, this stabilizing element is citrate and in *AA'* it is an unknown compound. These compounds and their surrounding water network provide many indirect contacts stabilizing the *BC* and *AA'* interfaces. In fact, only two direct contacts between MavE molecules are present outside this central, solvent-filled region. One of these is a hydrogen bond between the amide nitrogen of Ser36 and a terminal guanidinium nitrogen of Arg85. Since Ser36 is one of three tag residues retained after TEV cleavage, this interaction must be nonbiological. The other

direct contact is also a hydrogen bond, this time between Asp171 and Ser84. The limited direct interactions between MavE molecules, in conjunction with their monomeric state in solution, suggest that *BC* and *AA'* interactions do not represent a natural state of this protein in the cell.

Another interaction of ~283 $Å^2$ is present between molecules *A* and *C*. This interface is formed predominantly by helix B of chain *A* and the D-E loop of chain *C*. The only direct interaction between *A* and *C* takes the form of a hydrogen bond between O^{δ} of Asp146 in chain *C* and $N^{\delta 1}$ of His69 in chain *A*. An additional contact mediated by a water molecule exists between Gln72 and Ser153 of chains *A* and *C*, respectively. This interaction angles chain *A* relative to *C*, such that contacts formed between rows of *AA* and *BC* dimers are weaker than those between the latter and itself. Consequently, our crystal exhibits layers of *AA* dimers sandwiched between double-layers of *BC* dimers, when viewed along the a-axis.

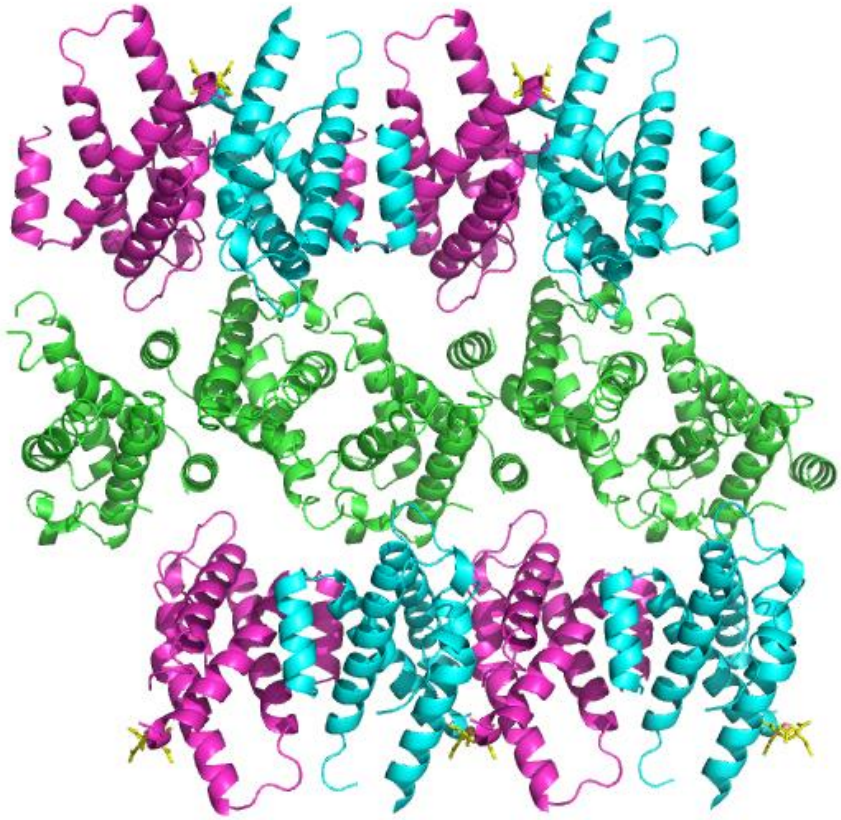

Figure 6. 4 Packing of MavE(39-172) molecules in the crystal. *AA'* molecules (green) are sandwiched between *BC* molecules (magenta and cyan) to produce a set of repeating layers, with each one rotated approximately 90 ° relative to its neighbor.

6.3.3. Comparison with the grass pollen allergen, Phlp5b

To gain insight into MavE function, we searched for proteins with a similar fold. To this end, we submitted MavE(39-172) coordinates to the DALI server. The translation initiation factor eIF2B (α-subunit) returned the highest Z-score (PDB id 5B04, Z-score = 7.4), however, MavE aligned with only a small segment of this protein. We searched the literature on eIF2B to understand the functional relevance of its α subunit. Structural studies on eIF2B provided insight into its role in translation. Briefly, eIF2B is a multiprotein complex that binds and deactivates another complex called eIF2. In its GTP-bound form, eIF2 initiates translation by delivering methionyl-tRNA to the ribosome. Recent work has shown that phosphorylated eIF2B binds in a groove between the α and β subunits to facilitate the exchange of GTP for GDP on eIF2 (Kashiwagi et al., 2016). This exchange allows translation to proceed by promoting the dissociation of eIF2 from the ribosome. Since MavE aligns with eIF2Bα, it may mediate an analogous protein-protein interaction to that seen between eIF2 and eIF2B. Closer inspection of the MavE-eIF2Bα alignment shows that helix B and its C-terminal loop have no counterparts in eIF2Bα. If these features were present in eIF2Bα, they would occlude a significant portion of the eIF2 binding interface. Specifically, Gln63, Gln87, His95 and Val97 in eIF2Bα are obscured by the loop connecting helix B to C in MavE. Thus, the structural similarity to eIF2Bα did not lead us to a functional understanding of MavE.

Two other proteins with a similar fold to MavE are the methylthioribose-1-phosphate isomerase (PDB id 6A34, Z-score = 6.9) and the ribose-1,5-bisphosphate isomerase (PDB id 5YFJ, Z-score = 6.8). Both proteins harbor an N-terminal helical bundle and a C-terminal catalytic domain. MavE aligns with a four-helix bundle at the N-terminus of each. These bundles have the same topology as MavE, with the second and fourth helix (violet and orange in Figure 6.5) converging at one end to form a V-shape that is interlocked with another inverted V from the first and third helices (red and green in Figure 6.5). Despite sharing this general conformation, substantial deviations can be seen in the orientation of certain helical axes. Specifically, helices A and E of MavE align with the ribose-1,5-bisphosphate isomerase, whereas helices A and C align with the methylthioribose-1-phosphate isomerase. The poor alignment of remaining helices may suggest weak conservation of the helical bundle architecture, likely arising from differences in the interhelical loop lengths.

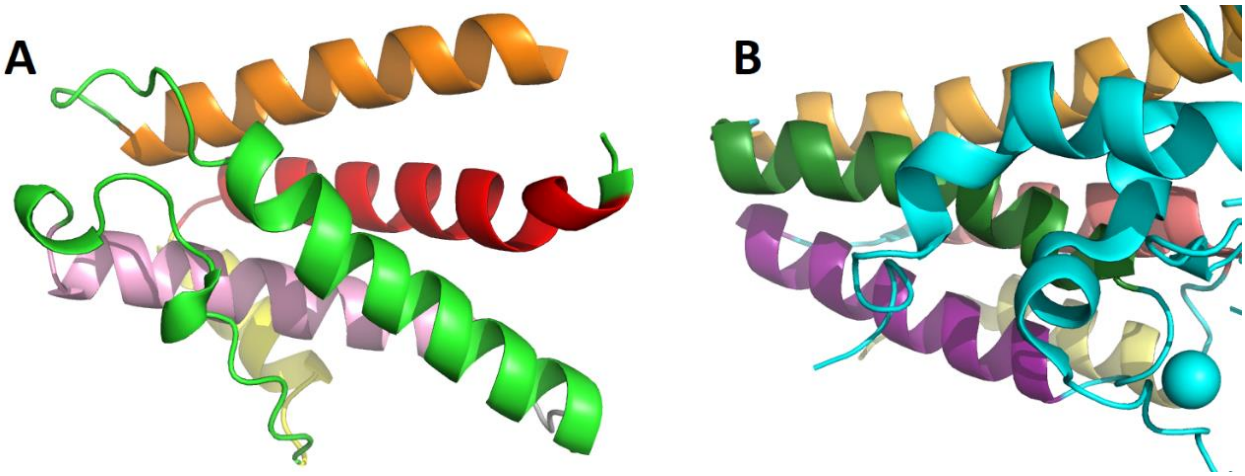

Figure 6. 5 MavE and ribose-1,5-bisphosphate isomerase adopt similar conformations in their 4-helix bundles.(A) Helices B and D (pink and orange) of MavE converge to form a < shape, which is interlocked with a > formed by helices A and C (red and green). (B) Ribose-1,5-bisphosphate has a similar organization of helices to MavE (related helices are coloured in similar shades.

Helix B in MavE has no counterpart in either methylthioribose-1-phosphate- or ribose-1,5-bisphosphate isomerase. However, both isomerases have N-terminal extensions that do not exist in MavE. The ribose-1-5-bisphosphate isomerase has a helical segment that is similar in length to helix B but does not align with it. Structural studies on the methylthioribose-1-phosphate isomerase show an N-terminal β-sheet and neighboring helix that shields the active site from solvent. These structural features are present on the same face of the four-helix bundle as helix B. It is therefore possible that the helical bundle represents a stable core domain upon which functional elements are grafted.

The most interesting result in our search for homologues was the similarity between MavE and the grass pollen allergen, Phlp 5b (PDB id 1L3P, Z-score = 7.1). Phlp 5b represents a core domain of the allergen comprised of a four-helix bundle (Rajashankar et al., 2002). Helices A, C, D and E of MavE(39-172) align well with the Phlp 5b core domain (Figure 6.5). This superposition is consistent with that seen in eIF2Bα and the two ribose-phosphate isomerases described above. That is, helix B has no counterpart in any of these homologues. Phlp 5b contains two 35-residue helix-turn-helix motifs sharing 37% sequence identity. These motifs adopt remarkably similar helix termination and chain reversal strategies. As such, Phlp 5b was described as a twinned two-helix bundle. The corresponding helix-turn-helix motifs in MavE

also share several residues, with some stabilizing the core architecture. Based on these observations, MavE can be said to have a similar core domain to that of Phlp 5b.

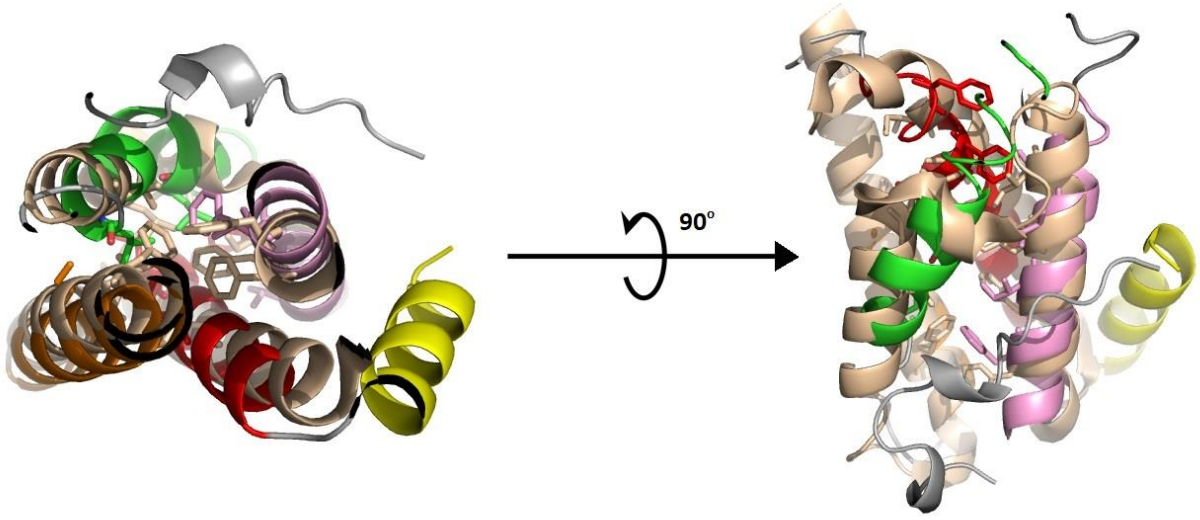

Figure 6. 6 Structural overlay of MavE(39-172) and the grass pollen allergen, Phlp 5b.MavE is coloured as in Figure 6.1. Phlp 5b is coloured wheat. Hydrophobic core residues stabilizing the twinned two-helix bundle are shown as sticks. Note that helix B (yellow) of MavE has no counterpart in Phlp 5b.

Despite being alike in their overall fold, the core domains of MavE and Phlp 5b share limited sequence identity. Some differences in sequence arise from the structural features distinguishing these proteins. Specifically, helices A and C are separated by helix B and an extended loop in MavE, while only a tight turn is present at this region in Phlp 5b. Inspection of the N-terminal motif in Phlp 5b shows that chain reversal is facilitated by a proline residue within this loop. This method of chain reversal is repeated in the C-terminal motif, where another proline is present on the loop connecting helices D and E. MavE contains such a proline residue only in the C-terminal motif (Figure 6.7). It is possible that the evolutionary pressure to retain proline at the N-terminal motif has been relaxed by the extended peptide connecting helices A and C.

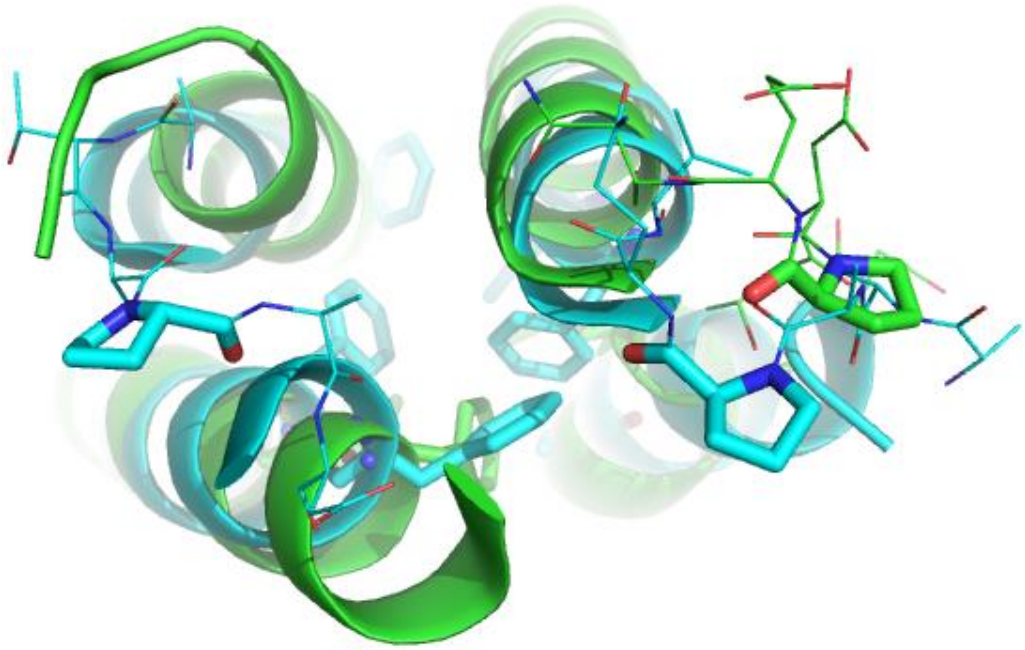

Figure 6. 7 Overlay of MavE(39-172) (green) and Phlp 5b (cyan) showing proline residues that facilitate chain reversal. Nearby loop regions are depicted as lines for clarity. Prolines are shown as sticks. Note the absence of a corresponding proline residue for MavE in the N-terminal motif (left).

Another unique feature in MavE is the presence of arginine residues at the beginning of each motif. These residues stabilize the core domain by mediating interactions between the two helix-turn-helix motifs. In contrast to Phlp 5b, which has isoleucine contributing to the hydrophobic core at this position, MavE uses water-mediated hydrogen bond networks to facilitate the contact of helix A with E and helix D with C.

Although Phlp 5b and MavE share a hydrophobic center, the contributing residues are different between these proteins. The hydrophobic center of Phlp 5b is comprised of positions 1, 4, 8, 24, 31 and 35 within each twinned motif. Conversely, MavE generates a hydrophobic center using positions 4, 8, 11, 15, 24 and 28 of the N-terminal segment, and positions 4, 8, 12 and 32 of the C-terminal segment. Differences between N- and C-terminal motifs in MavE may arise from asymmetry in the evolutionary pressures experienced by these regions of the protein. A possible source of this discrepancy is helix B and the residues connecting it to C, although it is unclear how this would affect the C-terminal motif.

6.3.4. Regions of functional significance in MavE

Helix B is the only secondary structure element in our MavE model that has no counterpart in Phlp 5b (Figure 6.4). Inspection of residues in the extended loop connecting helix B to C (Figure 6.1, grey region) reveals an NPxY motif. This sequence has been found to promote sorting or endocytosis by interacting with phosphotyrosine-binding (PTB) domain containing scaffold proteins (Stolt and Bock, 2006). Some proteins containing the NPxY motif are directed to the basolateral membrane (Stoops and Caplan, 2014). In other cases, a C-terminal NPxY motif allows low-density lipoproteins to be taken up by clathrin-mediated endocytosis (Chen et al., 1990). Recent studies have shown the NPxY motif to have divided roles for the organic anion transporting polypeptide (OATP1B1), with Asn and Pro residues promoting exit from the Golgi apparatus and Tyr playing a role in protein stability (Wang et al., 2019). These varied functions make the NPxY motif in MavE an attractive sequence to investigate; by mediating interactions with adaptor proteins that participate in vesicular trafficking, this motif may drive the LCV away from lysosomes. The electron density map surrounding the NPxY motif is poorly resolved in our structure, suggesting conformational flexibility in this region of the protein. The importance of the NPxY motif is further highlighted by its presence in many other species of *Legionella*. Considering the role of NPxY motifs in binding phosphotyrosine-binding domains (PTBs), it is possible that the twinned two-helix bundle acts as a scaffold to display this binding element to prospective partners.

In addition to the NPxY motif, MavE contains a set of residues along helices B and C that may act as a catalytic triad. Intriguingly, helix B contains two histidines, His68 and His69, and two aspartates, Asp64 and Asp65. These sidechains are in close proximity to Ser102 located on helix C. In particular, His68 could form a hydrogen bond to Ser102 upon rotation of the sidechain and at the same time come into proximity of the sidechains of Asp64 and/or Asp65. This Ser-His-Asp arrangement is highly reminiscent of the catalytic triad observed in peptidases. The residues of the putative peptidase triad are conserved in all *Legionella pneumophila* sequences but are not fully present in other *Legionella* species.

In summary, our MavE model contains a twinned two-helix core domain that is similar in fold to the grass pollen allergen, Phlp 5b. The repeated helix-turn-helix motifs comprising this core domain share limited sequence identity with those of Phlp 5b. These differences in motif

sequence give rise to alternative stabilization strategies, such as the use of arginine residues to facilitate intra-repeat interactions in MavE. Furthermore, MavE deviates structurally from Phlp 5b at the connection of helices A and C. Specifically, helix B and the loop connecting it to helix C is absent in Phlp 5b.

6.4. Discussion

MavE is a *Legionella pneumophila* effector protein secreted by the Icm/Dot Type IVB Secretion System. It has been called a meta-effector because it directly binds and inhibits the SNARE mimic, LegC7. We initially attempted to reproduce this interaction *in vitro* using pull-down assays. Briefly, purified His_6-LegC7(142-425) and GST-MavE(39-172) was added to either Co^{2+} and glutathione resin. Proteins were competitively eluted from the column using either imidazole or glutathione. This pull-down experiment was repeated using His_6-MavE(39-172) and GST-LegC7(142-425). Unfortunately, neither assay implied an interaction between LegC7 and MavE. This absence of binding was supported by the fact that His_6-LegC7(142-425) had no effect on the size exclusion profile of His_6-MavE(39-172). The biological significance of the MavE-LegC7 interaction has yet to be confirmed, however, and no function for MavE has been stated in the literature. Herein, we report the crystal structure of MavE(39-172) solved to 1.8 Å resolution. This construct lacks the C-terminal transmembrane region and some N-terminal residues erroneously included in the Uniprot annotation. Our MavE model contains a core domain reminiscent of the grass pollen allergen, Phlp 5b. The core domain of MavE is distinguished from the allergen by helix B and a loop connecting it to helix B. These structural elements may represent a region of functional significance for MavE, as their absence in Phlp 5b suggests they are dispensable for protein stability.

The twinned two-helix bundle core domain of MavE contains strikingly different motif sequences than those of Phlp 5b. This may be an example of evolution converging on a common fold. Although the reason for this convergence remains unclear, the literature on Phlp 5b suggests that structural stability is a driving factor. That is, the Phlp 5b core domain is resistant to proteolysis and retains allergenicity. The stability of the twinned two-helix bundle

can be attributed to its hydrophobic core and conserved residues promoting intramolecular contacts.

A characteristic property of the twinned two-helix core domain is proteolytic resistance. As such, MavE may have evolved this domain to act as a robust scaffold for displaying functional elements to binding partners. The fact that Phlp 5b retains allergenicity can further rationalize this claim, as *L. pneumophila* effectors must resist innate immune defences. This stability could be essential for effectors performing a critical function in virulence, especially those having no redundant counterpart.

Structural features outside the MavE core domain may be of functional significance to this protein. The loop connecting helices B and C is the longest in our model and contains a region of conformational heterogeneity, as evidenced by high B-factors and the absence of clear electron density. Intriguingly, the B-C loop harbours an NPxY sequence, which is the canonical phosphotyrosine-binding (PTB)-domain interacting motif. The presence of this motif on a flexible, solvent exposed loop in MavE may suggest a role for this protein in recruiting PTB-domains, which predominantly exist on adaptor or scaffold proteins (Uhlik et al, 2005).

Table 6. 1 Data collection and refinement statistics of MavE(39-172)

Data collection	
Space group	C121
Unit cell (Å)	199.5, 42.1, 46.3
Resolution (Å)	49.85 – 1.59 (1.6 – 1.59)
Total reflections	313021 (20373)
Unique reflections	80803 (5639)
Completeness (%)	80.8 (34.8)
Redundancy	3.87 (3.6)
Mean I/σ(I)	12.19 (1.11)
R_{merge} (%)	7.8 (127.8)
CC1/2	99.9 (59.3)
Refinement	
Resolution (Å)	46.27 – 1.65 (1.71 – 1.65)
No. of reflections	40116 (1946)
R_{work} / R_{free} (%)	19.7 / 23.7
No. of atoms / waters	3496 / 252
RMSD from ideal values	
Bond lengths (Å)	0.01
Bond angles (°)	1.07
Ramachandran plot (%)	
Favored	98.47
Allowed	1.28
Outlier	0.26
PDB ID	6PIR

7. Crystal Structure of the *Legionella* Effector MavL

7.1. Literature Review for MavL

MavL (*lpg2526*) is an *L. pneumophila* effector identified in the same manner as MavE (see section 2.4). Fusion of the MavL C-terminus to SidCΔ100 produces a hybrid protein capable of translocating through the T4BSS. Since SidCΔ100 is translocation deficient, MavL may harbor a signal sequence recognized by the T4BSS. A common motif present at the C-terminus of translocated effectors is a glutamate-rich stretch called the E-block. These acidic residues were later found to promote efficient translocation through the T4BSS by an interaction with DotM (see section 2.2.2.6) (Huang et al., 2011). Inspection of the MavL C-terminus reveals a potential E-block motif consisting of three consecutive glutamate residues, which may be responsible for rescuing the translocation deficit present in SidCΔ100.

MavL export has also been found to depend on the Lsp type II secretion system (T2SS), which relies on the Tat- or Sec- pathway to transport effectors across the inner membrane (DebRoy et al., 2006). Recruitment to either of these pathways requires an N-terminal signal sequence. This suggests that MavL contains an N-terminal sequence targeting it to the T2SS and a C-terminal (E-block) motif driving export by the T4BSS. Once in the periplasm, substrates of the T2SS interact with an outer membrane protein called secretin, which delivers them from the bacterial cell. Substrates of the T2SS are targeted to the LCV lumen, while those of the T4BSS are exported into the host cytoplasm (DebRoy et al., 2006). MavL may therefore play unique roles within and outside of the LCV during infection.

Secondary structure predictions using PsiPred show MavL to be predominantly α-helical with several short β-strands interspersed throughout. PsiBLAST on MavL returns several homologues with start sites corresponding to residue M41. As with MavE, this finding may point to an inaccurate start site annotation for this protein within the Uniprot database (https://www.uniprot.org/). Alternatively, it is possible that MavL harbors an N-terminal extension distinguishing it from known homologues.

MavL homologues are present in all known *L. pneumophila* strains but not in *L. longbeachae* (Gomez-Valero et al., 2011). This is surprising because both organisms produce

severe pneumonia in patients with weakened immune systems. The presence of MavL-like effectors in all *L. pneumophila* strains implies a role for this protein in virulence, while its absence in *L. longbeachae* suggests otherwise. It is possible that MavL acts redundantly with another *L. pneumophila* effector that has homologues in *L. longbeachae*. Indeed, *L. pneumophila* encodes over 330 effectors and this large number arises, in part, from functional redundancy. This phenomenon may be a consequence of multiple hosts in the environment; if *L. pneumophila* requires a unique complement of effectors to infect each host, expansion of the effector repertoire is required to broaden virulence (Ensminger, 2016). The significance of MavL to the pathogenicity of *L. pneumophila* has yet to be investigated.

Elizabeth L. Hartland and Shivani Pasricha have confirmed the Icm/Dot-dependent translocation of MavL into THP-1 macrophages using a fluorescence-based TEM β-lactamase assay. These researchers also found HA-tagged MavL to be distributed throughout the cytoplasm after being opsonized by Fcγ-HEK293 cells. Such findings suggest that MavL is exported to the cytoplasm and is not directed to membranous compartments of the host cell. Bioinformatic analysis using the Phyre2 server showed that 15% of the MavL sequence is homologous to poly-(ADP-ribose) polymerase (PARP) macrodomains. PARPs extend ADP-ribose chains on target proteins in eukaryotes. As such, MavL may interfere with the eukaryotic posttranslational machinery (unpublished data).

Another important discovery by Hartland and Pasricha is the direct interaction between MavL and the E2 ubiquitin conjugating enzyme, Ube2q1/Nice5. The binding event was initially identified by yeast two-hybrid and then in mammalian cells by co-immunoprecipitation (unpublished data). This finding implicates MavL in the ubiquitination pathway, although it remains unclear how modulation of this pathway occurs. Several *L. pneumophila* effectors act as E3-ubiquitin ligase mimics, which catalyze the non-canonical ubiquitination of host proteins. MavL may therefore prevent the interaction of Ube2q1 with host E3 enzymes and promote deviation of the ubiquitination pathway at the level of ligation. This deviation could result in the ubiquitination of host proteins that are not otherwise destined for degradation. Alternatively, MavL may chemically modify ubiquitin to render it inaccessible to host E3s. Further experiments are needed to deduce the role of MavL in ubiquitination.

L. pneumophila encodes an effector called Lem26 that causes a growth defect when overexpressed in yeast cells. Surprisingly, co-expression of MavL with Lem26 significantly reduces this defect, suggesting that Lem26 is inhibited by MavL. Interaction studies by yeast two-hybrid and co-immunoprecipitation did not show a direct interaction between these proteins, however (Urbanus et al., 2016). Although Lem26 remains structurally and functionally uncharacterized, a quick BLASTp search predicts the presence of Vip2- and SidE-domains at the N- and C-terminus of the protein, respectively. Vip2 domains typically play a role in ADP-ribosylation, while SidE domains have phosphodiesterase activity. This domain organization is reminiscent of SidE-family effectors, which promote the non-canonical ubiquitination of host proteins in an NAD^+-dependent process. The mechanism by which MavL impairs the yeast growth defect caused by Lem26 remains unknown.

7.2. Experimental Procedures for the MavL Project

7.2.1. Cloning of recombinant MavL

The gene encoding MavL (*lpg2526*) was cloned into pMCSG7 and pRL652 vectors by ligation independent cloning (LIC), as described in sections 4.2.1, 4.2.2 and 4.3.1. Initially, a construct consisting of residues 2-444 was designed. This construct expressed and purified poorly, compelling us to explore alternative constructs. As with MavE (section 4.3.1), the PsiBLAST algorithm indicated a start site mis-annotation for MavL; most homologues had start sites corresponding to the MavL residue M41. Two constructs of MavL were designed to investigate this putative start site, namely MavL(42-435) and MavL(42-388). Inserts corresponding to these sequences were amplified from the MavL(2-444) plasmid and inserted into pMSCG7 and pRL652 by LIC.

7.2.2. Protein expression and purification

MavL(42-435) and MavL(42-388) in pMCSG7 were first transformed into BL21(DE3)pLysS. Protein expression and purification was then carried out as described for LpnE, and OCRL in section 4.2.3. Limited proteolysis on MavL(42-435) and MavL(42-388) showed the latter construct to have greater susceptibility to degradation than the former. Furthermore, crystals were obtained only for a dimer species of MavL(42-435). This longer MavL construct was therefore used in all downstream structural and ligand-binding studies.

Purification of MavL(42-435) on a Biorad SEC650 or GE S200 column allowed for the separation of dimer from monomer species, which were then concentrated to 15 – 20 mg/ml for crystallization trials. Initially, MavL(42-435) eluted from size exclusion predominantly in the monomeric form. The dimer species became more prominent after a day at 4 °C. By the third day, the dimer peak was larger than that of the monomer. At this point, further incubation of MavL(42-435) at 4 °C no longer increased the presence of the dimer. Re-running monomeric MavL on gel filtration regenerated some dimer and vice versa. These findings show that monomer and dimer species of MavL(42-435) are in equilibrium, with the monomer favored initially.

Intriguingly, only the MavL(42-435) dimer was found to crystallize. This observation comes in conflict with the previously reported equilibrium between monomer and dimer species. To account for this, it is possible that mixing purified MavL(42-435) with the crystallization solution stabilized the dimer. If so, rapid screening of the dimer after separation would have been essential to the crystallization of MavL(42-435). Consistent with this notion, screening of MavL was always carried out immediately after purification and any protein that could not be screened was flash-frozen in liquid nitrogen. In the few cases where protein was left at 4 °C, it gradually lost its crystallizeability.

No phasing model for MavL(42-435) was present in the Protein Data Bank (PDB). We therefore opted to produce a seleno-methionine derivative for phasing by single anomalous dispersion (SAD). Inhibition of methionine biosynthesis was initially attempted in BL21(DE3)pLysS cells harboring the MavL(42-435) plasmid in pMCSG7. We found that growth of these cells in minimal media was prohibitively impaired. To overcome this, the MavL(42-435) plasmid was transformed into auxotrophic B834 *E. coli* competent cells. These

transformants grew well in minimal media supplemented with methionine, but not without it. Induction of protein expression was carried out as previously, using 500 µL IPTG and reducing the temperature to 18 °C. Importantly, the culture was resuspended in minimal media lacking methionine immediately prior to induction, at which point seleno-methionine was added to the culture. In this way, the incorporation of seleno-methionine into recombinant protein was ensured. Purification of seleno-methionine (SeMet) MavL(42-435) was carried out as described for native MavL(42-435).

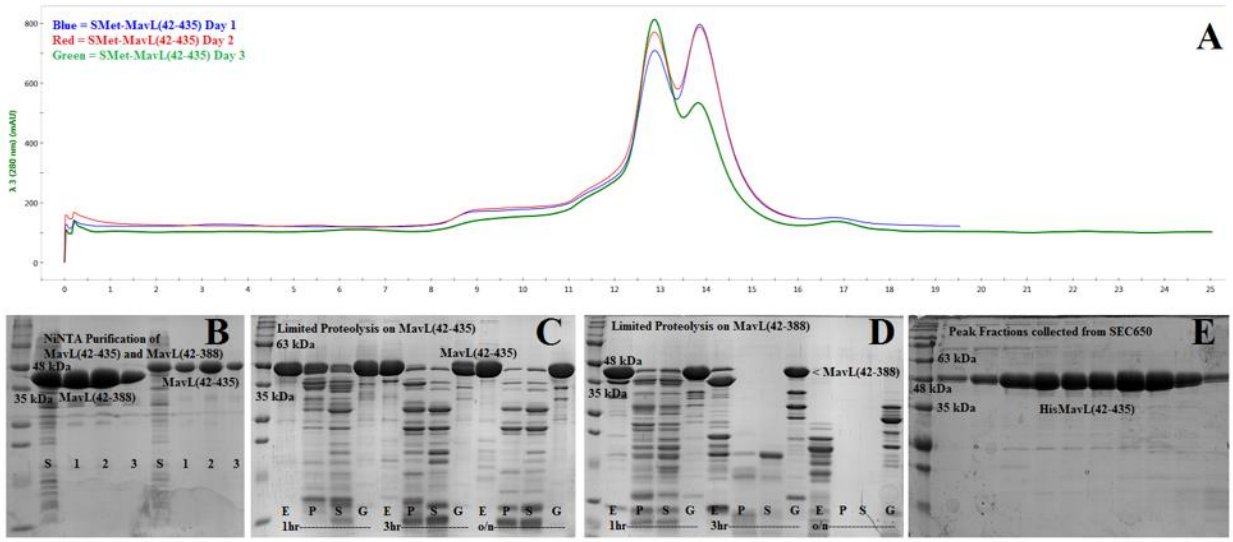

Figure 7. 1 Purification of His-MavL(42-435) and His-MavL(42-388). (A) Size exclusion chromatograms of SMet-MavL(42-435) showing the transition from monomer to dimer species. The early stage run (blue) shows more monomer than dimer. Later (red), nearly the same amount of monomer and dimer are present. After three days (green), the dimer predominates. (B) Purification of His-MavL(42-435) and MavL(42-388) on NiNTA. S = soluble, 1 – 3 = elutions 1 – 3 in 15 mM Tris, pH 8.0, 50 mM NaCl, 100 mM imidazole. (C) and (D) Limited Proteolysis on His-MavL(42-435) and His-MavL(42-388), respectively. E = elastase, P = proteinase K, S = subtilisin and G = GluC. (E) Peak fractions collected from the size exclusion run of SMet-MavL(42-435) (depicted in A).

7.2.3. Crystallization of MavL(42-435)

Initial crystals for MavL(42-435) were obtained from Crystal Screen HT (Hampton Research, Aliso Viejo, CA). Specifically, conditions A4 (0.1 M TRIS hydrochloride pH 8.5, 2.0 M Ammonium sulfate) and D3 (0.1 M HEPES sodium pH 7.5, 2% v/v Polyethylene glycol

400, 2.0 M ammonium sulfate) gave poorly diffracting needle clusters in under three days of growth. Extensive efforts to optimize the morphology of these crystals using additive- and grid-screens were met with little success. A few months later, an improved crystal form of MavL(42-435) was discovered in Crystal Screen condition B3 (0.2 M ammonium sulfate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% w/v polyethylene glycol 8,000). These slower growing crystals diffracted better than the initial needle-clusters. After optimization, the best crystals were obtained at 20 °C in drops containing 1 µL protein in 15 mM Tris-HCl, pH 8.0, 50 mM NaCl mixed with 1 µL reservoir comprised of 0.2 M ammonium sulfate, 0.1 M citrate pH 6.0, 22% w/v polyethylene glycol 8,000. Both native and seleno-methionine derivative MavL produced optimal crystals in this condition.

7.2.4. Data collection and structure determination

Seleno-methionine derivative MavL crystals were transferred into 1 µL mother liquor containing 25% (v/v) glycerol. Diffraction data was collected to 2.65 Å at the Canadian Macromolecular Crystallography Facility (CMCF) 08ID beamline, Canadian Light Source (CLS), using a Pilatus3 S 6M detector (Grochulski et al., 2011). Integration and scaling were carried out using the XDS software package (Kabsch, 2010). Initial electron density maps were obtained in Phenix using the hybrid substructure search to determine heavy atom (selenium) sites followed by phaser EP to calculate phase estimates for all remaining structure factors. The autobuild script was then used to arrive at a suitable model for further refinement (Adams et al., 2011). Manual adjustments to this model were made in Coot (Shabalin et al., 2018) followed by computational refinement using phenix.refine (Adams et al., 2011).

7.2.5. Isothermal titration calorimetry of MavL(42-435) with ADP-ribose

Titrations were carried out using the Nano ITC instrument (TA Instruments, New Castle, DE) and analyzed with the NANOANALYZE software (TA Instruments, New Castle, DE, USA) as described in section 4.2.8. Briefly, 400 µM ADP-ribose was titrated into the calorimeter cell containing 100 µM HisMavL(42-435). Both ligand and protein were in a

buffer containing 15 mM Tris-HCl, pH 8.0 and 50 mM NaCl. The experiment was carried out at 20 °C (see section 4.5.2 for detailed explanation of ITC).

7.3. Results

7.3.1. MavL shares the fold of ADP-ribose binding macrodomains

To learn more about the biochemical function of MavL, the crystal structure of His-MavL(42-435) was solved by single anomalous dispersion (SAD) with density modification and refined to a resolution of 2.65 Å. There are three molecules in the asymmetric unit. They are nearly identical and superimpose with a root-mean-square deviation (rmsd) for Cαs of ~0.4 Å. The core of each molecule is formed by a highly curved and twisted, ten-stranded mixed β-sheet (Figure 7.2). Two α-helices line the concave side of the sheet and are nearly parallel to the β-strands, while long loops with embedded short α-helices cover the convex side of the sheet (Figure 7.2). Among them is a bundle of four short α-helices that are inserted between strands β1 and β2 and abate to one end of the β-sheet.

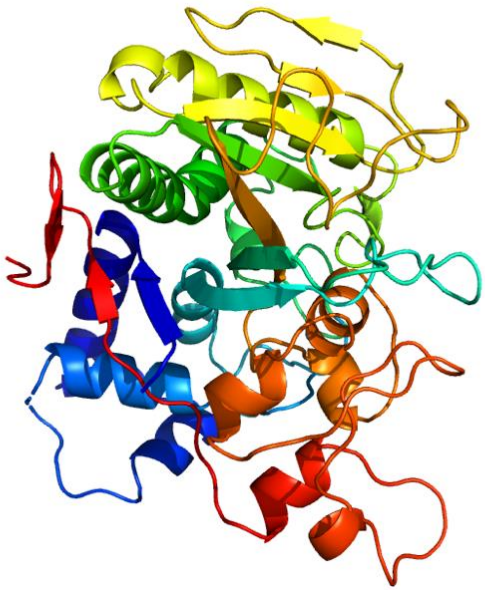

Figure 7. 2 Overall structure of MavL(42-435). Chain is coloured blue-green-red from N- to C-terminus. Note the α-helical bundles flanking each side of the central β-sheet.

A quest for proteins with similar fold conducted with the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/, (Holm & Rosenström, 2010) identified partial similarity to proteins containing the ADP-ribose binding macrodomain fold. The region of similarity encompasses residues 130-340 out of the 41-404 MavL sequence. The most compact ADPR-binding macrodomain contains ~200 residues and is composed of a seven-stranded mixed β-sheet with three long α-helices on one side covering the entire face of the β-sheet, and two α-helices lining the end of the β-sheet on the other side. Superposition of MavL and the smallest macrodomain from PARP14 (PDB ID 5QHT, PanDDA analysis group depsotision, to be published) shows good correspondence for seven β-strands, two of the three long α-helices on one side of the sheet and one short α-helix on the other side. However, while the arrangement of β-strands is conserved, their connectivity/topology differs between MavL and the classic ADPR-binding macrodomains. In MavL these seven strands are in order β1-β2-β7-β3-β4-β6-β5 while in the classic fold the strand sequence is β1-β2-β7-β6-β3-β5-β4. Thus, MavL folds in a similar three-dimensional arrangement but with a rearranged order of β-strands and with special conservation of only three helices.

The N- and C-terminal extension to the ADP-ribose binding macrodomain fold do not form separate domains but supplement the macrodomain fold by adding three strands to the central β-sheet and several α-helices. The question then arises if this modified macrodomain fold still recognizes ADP-ribose. Superposition of MavL with the ARTD8 macrodomain m1, m2 or m3 containing ADP-ribose shows that the substrate binding site of all these proteins has similar shape and the ADP-ribose fits in this superposition neatly into the potential MavL binding site. To experimentally test the binding of ADP-ribose to MavL we performed isothermal titration calorimetry (ITC) (Figure 7.4)). The binding was easily detected and the estimated binding constant Kd was 13μM.

The proteins containing ADP-ribose binding macrodomains are classified as readers, writers and erasers (Karlberg et al, 2013). Reader domains recognize and bind ADP-ribose modifications and are usually embedded within large multidomain proteins. Writers are components of ADP-ribosyl transferases that synthesize ADP-ribose from NAD+ and modify target proteins on a specific amino acid. Finally, erasers hydrolyse poly(ADP-ribose) at the glycosidic (1-2') linkage. We analysed in more details the structural similarity of MavL to

proteins with macrodomains from each of these three classes to identify which of these three functions is most likely for MavL.

Structural overlay of MavL with 5KIV, 3SIH and 3VFQ shows a common ADP-ribose binding pocket formed by five loops. Loops I and V interact with the adenine moiety of ADP-ribose, while loops II – IV bind the pyrophosphate and terminal ribose (ribose") (Figure 7.3). Extensive contacts with the ligand are made by loop IV, which harbors a conserved pyrophosphate binding motif (hGTGxh) where the threonine hydroxyl hydrogen bonds with the 3' oxygen of ribose'. The structurally corresponding motif in MavL, $I^{222}GTGCF^{227}$ and Thr224 could form the same hydrogen bond with the O3' of ribose.

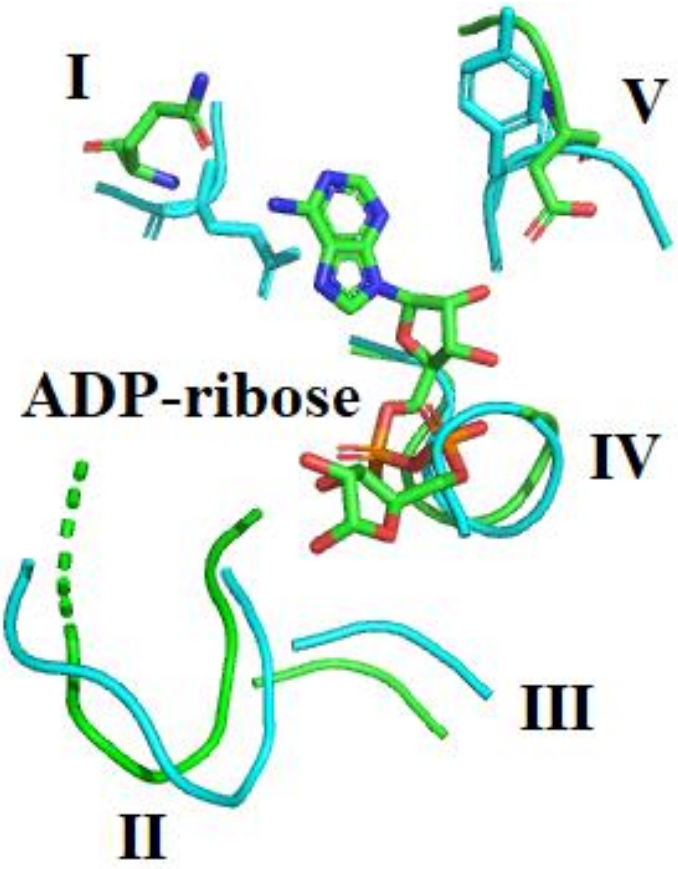

Figure 7. 3 Overlay of MavL(42-435) (cyan, solid) and 3VFQ (green, transparent). ADP-ribose from 3VFQ is depicted as green sticks and demarcates the putative ligand-binding pocket in MavL. Note the similar conformation of loops surrounding ADP-ribose.

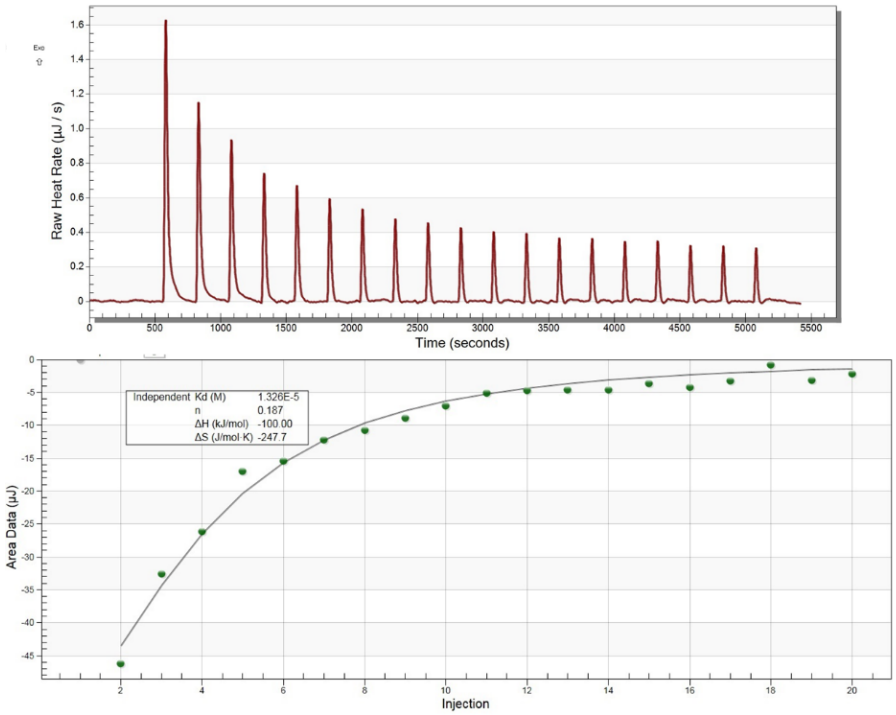

Figure 7. 4 Isothermal titration calorimetry (ITC) on His-MavL(42-435) and ADP-ribose. 400 µM ADP-ribose was titrated into 100 µM His-MavL(42-435). Top = Raw data on exothermic emissions for 19 successive injections (the first injection emitted negligible heat). Bottom = Binding curve obtained from the raw data.

All three ADP-ribose binding homologues (5KIV, 3SIH, 3VFQ) share a set of five α-helices flanking the substrate binding cleft. These helices are a scaffold upon which the ligand binding loops are grafted and numbered in the same way as the loops. Although MavL retains helices II – IV, it lacks helices I and V entirely. These latter helices give rise to the adenine binding loops, which are conformationally more variable than the pyrophosphate-binding counterparts (loops II – IV). This renders greater flexibility in loops I and V of MavL, as evidenced by conformational differences between homologues. Loop I of MavL shows the greatest deviation in our model, which could be attributed to the absence of bound ADP-ribose. The remaining helices of MavL, 5KIV, 3SIH and 3VFQ have no counterparts in the other models and likely play a structural role. In summary, the central β-sheet and helices II – IV forming the ligand binding cleft are conserved features in ADP-ribose binding macrodomains, likely due to their role in stabilizing the core architecture and promoting the interaction with the pyrophosphate of ADP-ribose.

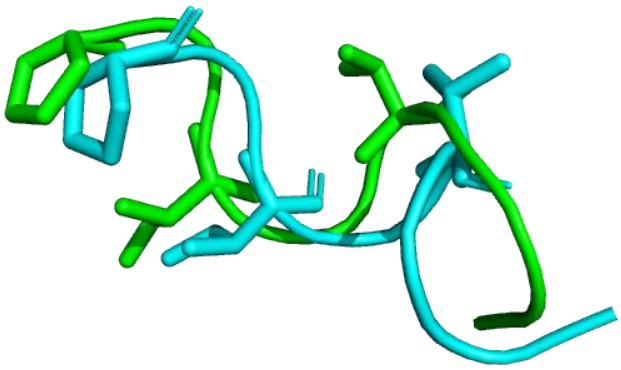

Figure 7. 5 Alignment of the pyrophosphate-binding loop IV in MavL and ADP-ribose binding homologue, 3VFQ. Note the similar organization of proline, isoleucine and threonine residues along each loop. Mav2 and 3VFQ are depicted in cyan and green, respectively. The threonine residue in MavL adopts a different conformation than that of 3VFQ and is not positioned for ligand binding.

The similarity of MavL to ADP-ribose binding proteins can lead to the identification of putative catalytic residues. For example, MavL may act as a poly-(ADP-ribose) polymerase (PARP) or poly-(ADP-ribose) glycohydrolase (PARG). Indeed, the MavL homologue 3VFQ is a PARP, while 3SIH is a PARG. Despite differing in their catalytic mechanism, the ligand binding loops of 3VFQ and 3SIH adopt remarkably similar conformations, with only loop V being distinct between them. A mechanism for the glycohydrolase activity of 3SIH has been proposed, where the side chain carboxyl of a glutamate (Glu115) on loop II is deprotonated by the glycosidic oxygen that links ribose moieties in the chain. This leads to the formation of a transient oxocarbenium intermediate that is neutralized by a water molecule. Prior to the reaction, a neighboring glutamate (Glu114) and phenylalanine (Phe227) residue orient the ribose moiety to facilitate cleavage (Slade et al., 2011). A cluster of three glycine residues interact with the pyrophosphate of a neighboring ADP-ribose moiety and are also essential for glycohydrolase activity in 3SIH (Patel et al., 2005). Together, these residues form a conserved GGG-X_{6-8}-QEE motif that characterizes PARG enzymes. In contrast to 3SIH, 3VFQ lacks acidic residues on loop II. Furthermore, Phe227 in 3SIH is replaced by a Leu1137 in 3VFQ and no cluster of glycine residues is present in the corresponding position of 3VFQ. Thus,

although ADP-ribose polymerases and glycohydrolases share structurally similar ligand binding pockets, they can be distinguished by the GGG-X_{6-8}-QEE motif. The ligand binding loops of MavL deviate slightly from the conformations observed in 3SIH and 3VFQ, with the largest differences being in loops I and V. Structural alignment of MavL with 3SIH reveals a pair of aspartate residues (Asp332, Asp333) occupying the same site as the catalytic glutamates in 3SIH (Figure 7.6). In addition, the phenylalanine (Phe227) residue in 3SIH (that positions the ribose to be cleaved) is paralleled by another Phe227 in MavL. This residue takes on different conformations in these proteins, with the sidechain interacting hydrophobically with the terminal ribose in 3SIH but not in MavL. The unique orientation of Phe227 in MavL suggests flexibility in loop IV, a property that is compatible with substrate binding. Thus, structural analysis of MavL points toward PARG activity, with Asp333 on loop II initiating the cleavage reaction and Phe227 stabilizing the ribose moiety. Further experimentation is needed to verify if such activity is truly present for MavL.

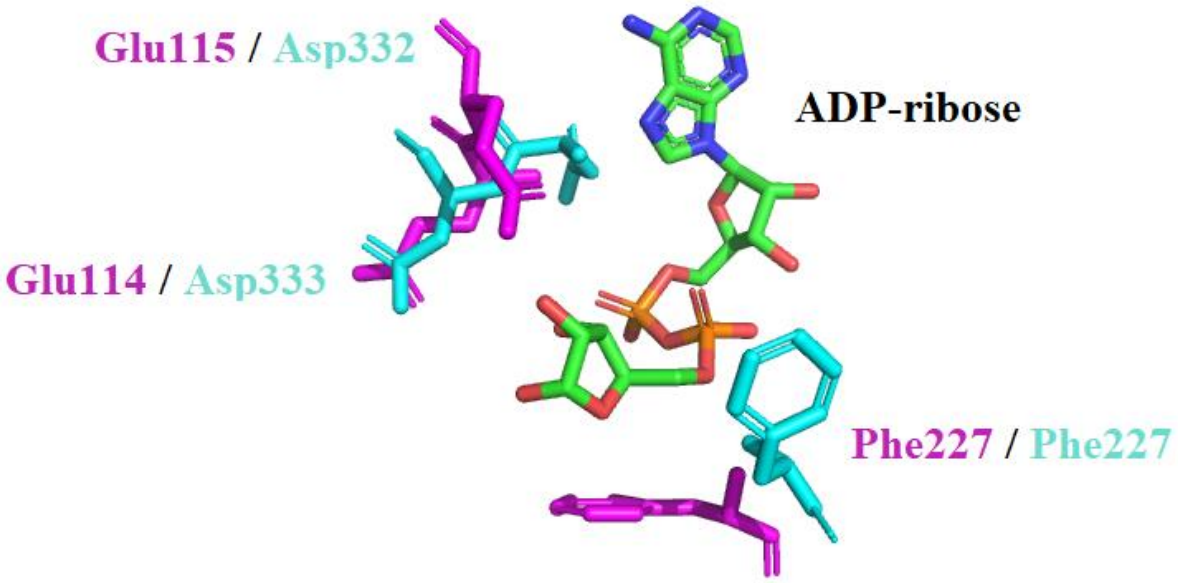

Figure 7. 6 Neighboring aspartate residues in MavL (cyan) occupy a similar position to the catalytic glutamates of 3SIH (magenta). Coordinates for ADP-ribose (green) are taken from the 3VFQ PDB. Note the similar orientation of Glu114 (3SIH) and Asp333 (MavL) sidechains, which coordinate the terminal ribose moiety. Conformational differences between remaining sidechains may reflect the flexibility needed to accommodate ligand.

7.3.2. Crystal contacts

Despite having three molecules in the asymmetric unit, the MavL lattice consists of stacked dimers that are rotated to form twisted columns. These dimers are comprised of molecules *A* and *B* or *C* and *C'* from a neighboring unit cell. The presence of dimers in the MavL crystal structure is consistent with its oligomeric state in solution. That is, only MavL fractions collected from the dimer peak on size exclusion could be crystallized. Each monomer is related to its counterpart by a twofold symmetry axis and forms a contact interface of 1067 $Å^2$. Strikingly, the putative ADP-ribose binding pockets are facing one another in the dimer, with several interacting residues arising from the C-terminal end of loops III, IV and V. Specifically, Tyr232 of loop IV hydrogen bonds with backbone atoms of Asp166 and Ser167 extending from loop III in the neighboring dimer. The methyl groups of Asp230 in both chains are near one another and contribute to the hydrophobic interface connecting the molecules. Two loops at the C-terminal end of each chain also stabilize the interaction. Another significant contribution to the dimer is a stacking contact between the aromatic rings of Trp357 in each chain (Figure 7.8). Perhaps the most convincing evidence for the biological relevance of dimeric MavL is a disulfide bond joining Cys226 residues in the *CC'* dimer. Although no disulfide is present in the *AB'* dimer, superposition of the *CC'* dimer reveals a similar orientation of Cys226 residues in both. It is possible that spontaneous reduction of the Cys226-Cys226 disulfide contributes to the monomer-dimer equilibrium observed for MavL by regenerating the monomer species. This hypothesis is consistent with the dynamic presence of MavL monomers and dimers in solution.

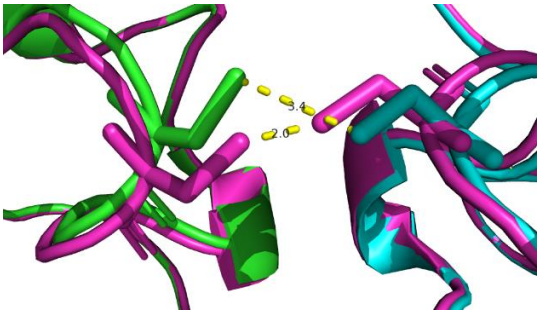

Figure 7. 7 Close-up of Cys226 residues in the CC' (magenta) and AB (green and cyan) MavL dimers. The presence of a disulfide between C and C' is indicated by the 2.0 Å distance between sulfur atoms. Slight conformational differences between Cys226 residues in the AB dimer separate their sulfur atoms by 3.5 Å, suggesting that these residues are nonbonded.

Intriguingly, electrostatic surface rendering shows a channel passing through the dimer interface that provides access to the putative ADP-ribose binding site (Figure 7.9). This is the only channel present in the dimer and may accommodate the binding of poly-(ADP-ribose) chains to facilitate reader or eraser activity. In addition, purified MavL gradually transitions from monomer to dimer species in solution, with the equilibrium favoring the dimer species. Taken together, these structural and biochemical insights suggest a biological significance for the dimeric state of MavL.

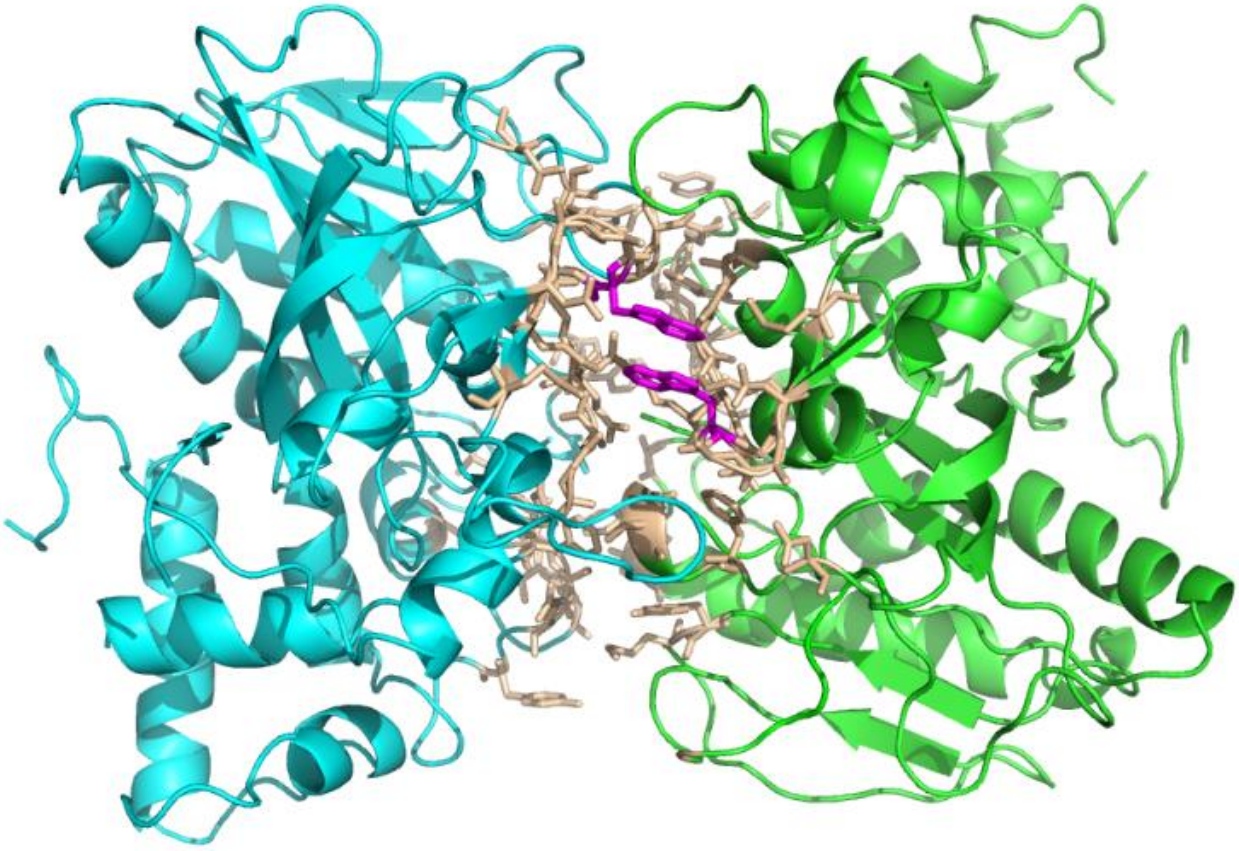

Figure 7. 8 The MavL dimer interface. Molecules A and B are depicted in green and cyan, respectively. Binding residues are represented as wheat coloured sticks and the C-terminal stacking interaction between W357 of each chain is shown in magenta.

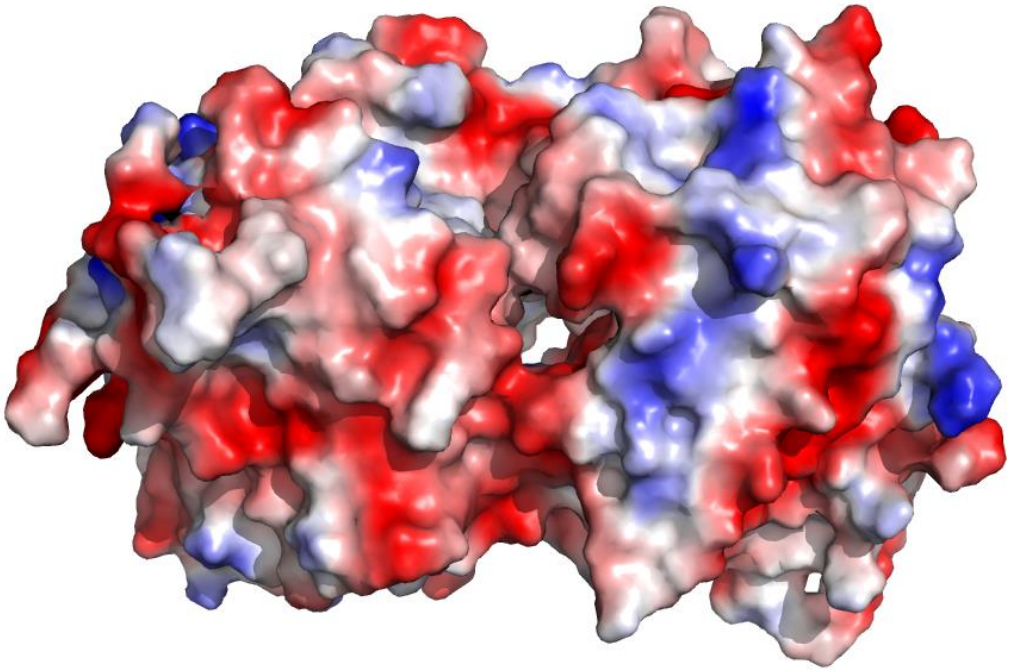

Figure 7. 9 Surface electrostatic rendering of the MavL dimer. A solvent filled channel passes through the dimer interface and exposes the active site of each monomer.

In contrast to MavL, all three ADP-ribose binding homologues (5KIV, 3SIH, 3VFQ) are monomeric. Structural alignment shows N-terminal α-helices in both 5KIV and 3SIH that protrude into the MavL dimer interface where C-terminal contacts would otherwise be made. Although no obvious steric prevention of dimerization is present in 3SIH, this model lacks structural analogs to the C-terminal binding loops and Y232 in loop IV of MavL. Our data shows a unique mode of dimerization for ADP-ribose binding proteins. The biological role of this interaction has yet to be evaluated.

7.3.3. MavL/Ube2q1 interactions

Our collaborators Shivani Pasricha and Elizabeth Hartland have identified an interaction between the MavL effector and the E2 ubiquitin conjugating enzyme, Ube2q1, by the yeast two-hybrid method and by co-immunoprecipitation. We aimed to reproduce this interaction *in vitro* using recombinant proteins. Two constructs of Ube2q1 were used for this purpose: (1) full-length Ube2q1(1-422) and (2) Ube2q1(247-414), which corresponds to a previously crystallized construct (PDB id 2QGX) (Sheng et al., 2012). The interaction of Ube2q1 with MavL was investigated using size exclusion chromatography. Initially, HisUbe2q1(1-422) and HisMavL(42-435) were purified on NiNTA, dialyzed in a running buffer (15 mM Tris, pH 8.0, 50 mM NaCl) and mixed in equimolar amounts. The putative complex was incubated for a minimum of 30 minutes on ice. When injected onto the column together, HisUbe2q1(1-422) and HisMavL(42-435) produced a chromatogram that looked like an overlay of the two profiles representing independent proteins (Figure 7.10). Although MavL(42-435) and Ube2q1(1-422) have similar molecular weights, the lack of a shifted peak was taken to indicate an absence of binding between these constructs. Next, the putative interaction of Ube2q1(247-414) with MavL(42-435) was explored by size exclusion chromatography. Ube2q1(247-414) was found to elute in a distinct peak that did not shift in the presence of MavL(42-435), suggesting that this truncated construct is also unable to interact with MavL (Figure 7.11).

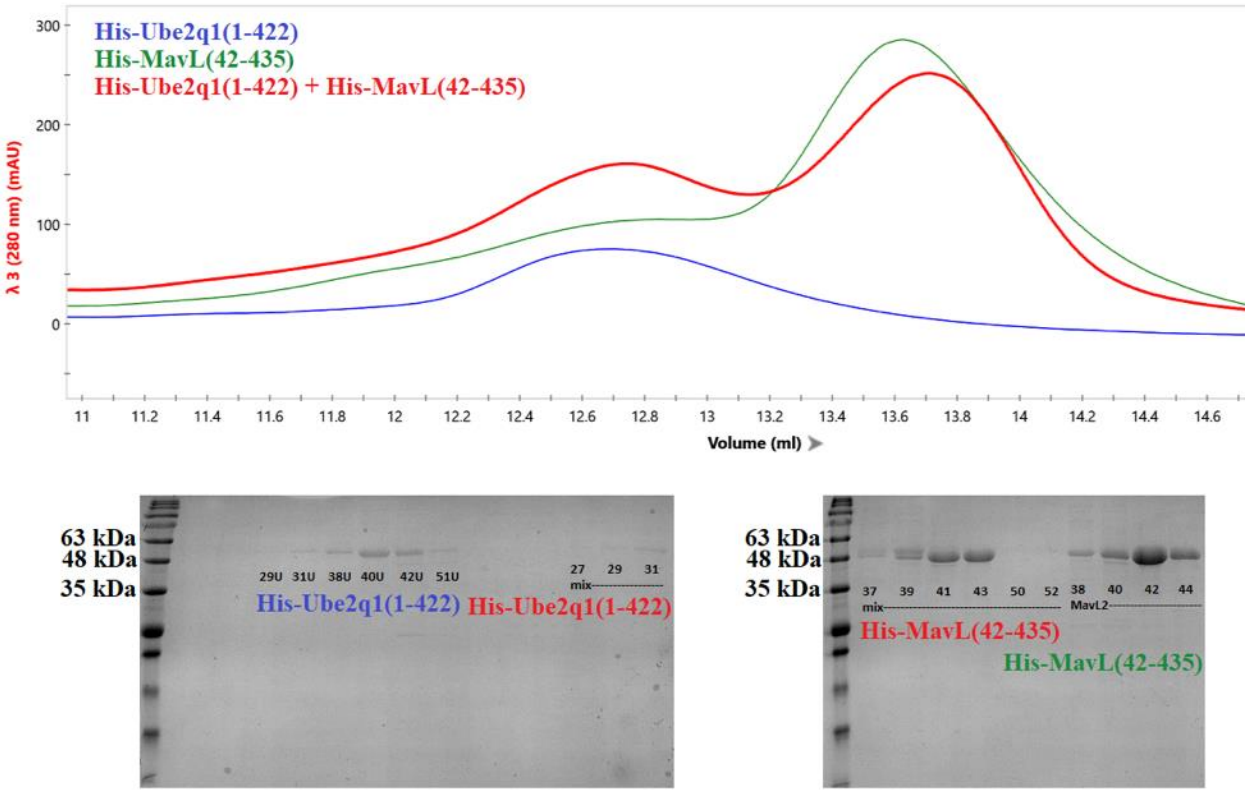

Figure 7. 10 Size exclusion chromatography on **His-Ube2q1(1-422) [Blue]**, **His-MavL(42-435) [Green]** and the **putative complex [Red]** using a Biorad SEC650 column. In the complex profile, His-Ube2q1(1-422) elutes slightly before His-MavL(42-435) in a peak that is unshifted relative to the independent protein. The His-MavL(42-435) peak is also unchanged by the presence of His-Ube2q1(1-422), indicating no interaction between these proteins. The SDS-PAGE analysis of each run is shown in the two gels below the chromatogram.

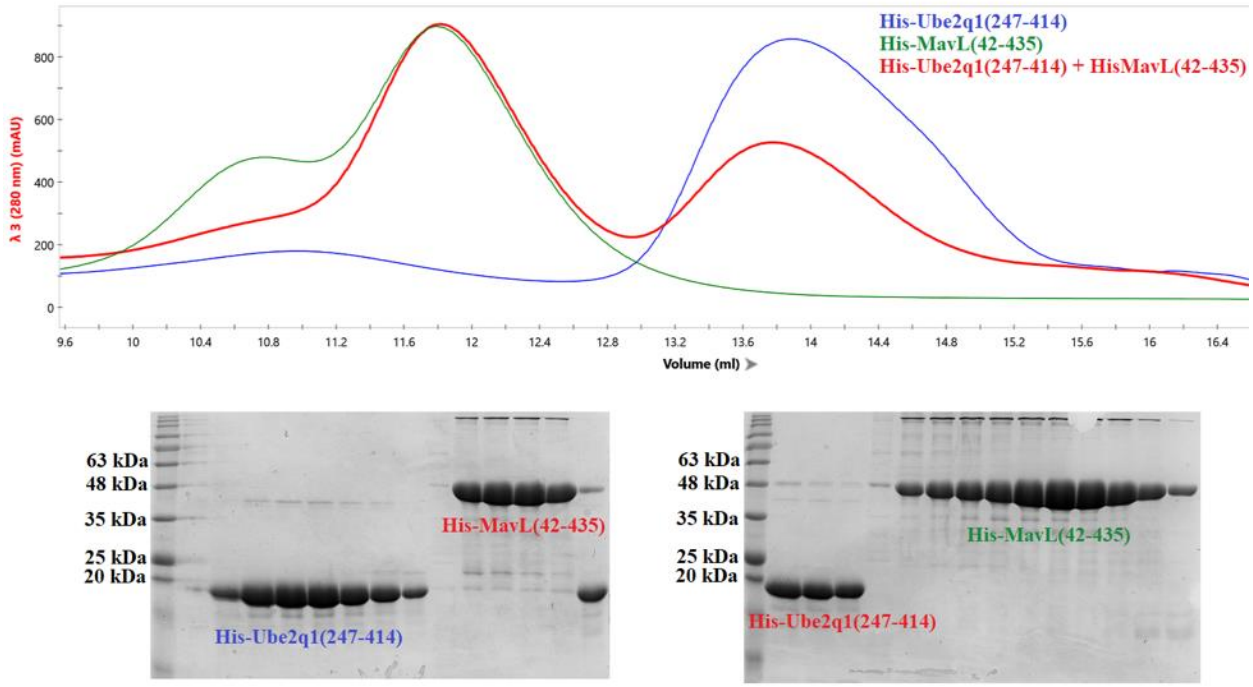

Figure 7. 11 Size exclusion chromatography (SEC) on **His-Ube2q1(247-414) (PDB id 2QGX) [Blue]**, **His-MavL(42-435) [Green]** and the **putative complex [red]**. In the complex profile, His-MavL(42-435) and His-Ube2q1(247-414) elute in distinct peaks that are unchanged relative to that seen for individual proteins. The SDS-PAGE analysis of each run is shown in the two gels below the chromatogram.

7.4. Discussion

The crystal structure of MavL(42-435) reveals a similar fold to ADP-ribose binding proteins, with potentially catalytic aspartate residues occupying the same position as glutamates characterizing the PARG motif (GGG-X_{6-8}-QEE). Aberrant ADP-ribosylation can have severe consequences for the host cell, including the inhibition of protein synthesis or overproduction of cAMP. Many species of pathogenic bacteria have exploited this powerful modification using secreted effectors that exhibit mono- or poly-ADP-ribosyltransferase (MAR/PAR) activity, leading to deleterious consequences for the host. Reversal of the PARP modification depends on enzymes with poly-(ADP-ribose) glycohydrolase (PARG) activity. Enzymes responsible for initiating or extending a poly-(ADP-ribose) chain on a protein are called writers, while those that hydrolyze ADP-ribose moieties from an existing chain are erasers. MAR- and PARylation is regulated by writers and erasers to mitigate the effect that would otherwise accompany unbridled modifications of this kind. If MavL does indeed have eraser activity, it could mitigate the effect of a bacterial or eukaryotic writer present in the host cell.

A third category of proteins, called readers, recognize ADP-ribosylated proteins and initiate protein-protein interactions that set signaling cascades into motion. Intriguingly, MavL shares structural homology with ARTD8 macrodomain 2 (ARTD8m2). ARTD8m2 distinguishes mono- from poly-ADP-ribosylated proteins and is classified as a reader. Since macrodomains recognize the entire ADP-ribose moiety at the site of attachment, they often form complexes with the modified protein. Mutations disrupting macrodomain function have been associated with various diseases, including cancer and neurodegeneration (Schuller et al., 2017). The similarity to ARTD8m2 may suggest reader activity for MavL. Alternatively, it is possible that eraser activity of MavL is tailored for mono-ADP-ribosylated proteins, although it remains unclear how mono- and poly-(ADP-ribosylated) proteins are distinguished by these macrodomains.

MavL was found to mitigate the growth defect observed in yeast cells expressing Lem26, another *L. pneumophila* effector containing an N-terminal Vip2 domain and a C-terminal SidE-domain. No interaction was observed between these proteins by yeast two-hybrid or (LUminescence based Mammalian IntERactome) LUMIER assays in HEK293 cells. These

findings suggest that MavL inhibits Lem26 by an indirect mechanism. VIPs (vegetative insecticidal proteins) are a pair of toxins found in *Bacillus cereus*, where VIP1 forms a multimeric pore in the membrane, allowing VIP2 to enter the host cell. VIP2 has been shown to ADP-ribosylate actin using a Vip2 domain that has since been reported in several other proteins (Han et al., 1999). Meanwhile, effectors belonging to the SidE family catalyze the noncanonical ubiquitination of host proteins in a process that does not depend on host cell machinery (i.e. E1 activating, E2 conjugating and E3 ubiquitin ligases). In the first step of this process, ubiquitin is ADP-ribosylated by a mono-ADP-ribosyltransferase (mART) domain at the N-terminus of SidE-family proteins. The modified ubiquitin is then transferred to a C-terminal phosphodiesterase (PDE), or SidE-domain, that renders phosphoribosyl-ubiquitin (PR-Ub), which is attached to a serine residue on the target protein. Surprisingly, the SidE domain of an effector called lpg1496 did not exhibit PDE activity against an array of nucleotides tested. This domain was found to bind ADP, however, suggesting that SidE domains do not exclusively cleave phosphodiester bonds in nucleotides. Although the function of Lem26 remains unknown, the presence of both Vip2 and SidE domains in this protein suggests a role in ADP-ribosylation. Thus, Lem26 may act as a bacterial writer protein. The indirect inhibition of Lem26 activity by MavL could be a result of opposing functions on the same target. That is, ADP-ribose moieties added by the Lem26 Vip2 domain are removed by MavL eraser activity. Alternatively, MavL might act as a reader of proteins modified by Vip2, initiating a signaling event that negatively regulates Lem26 activity.

Our collaborators, Shivani Pasricha and Elizabeth L. Hartland, found MavL to interact with the E2 ubiquitin conjugating enzyme, Ube2q1/Nice5, by yeast two-hybrid and co-immunoprecipitation experiments. We were unable to visualize this interaction *in vitro* by size exclusion chromatography with purified recombinant proteins. This may suggest a transient interaction between MavL and Ube2q1 or that some modification of either Ube2q1 or MavL is a prerequisite to binding. Considering the role of Ube2q1 in ubiquitination, it is possible that MavL only binds Ube2q1 in complex with ubiquitin. To explore this, we mutated the catalytic cysteine of Ube2q1 to serine and generated Ube2q1-ubiquitin complexes linked by an ester bond. We also used a construct of ubiquitin with its terminal glycine mutated to serine to generate a disulfide linked Ube2q1-ubiquitin complex. Neither of these complexes bound

MavL by pull-down or size exclusion chromatography. While these findings may imply that the MavL-Ube2q1 interaction is weak and transient, they could also suggest that additional modifications are required for binding. The structural studies presented here point to ADP-ribosylation as a likely recognition determinant for MavL. Since both Ube2q1 and ADP-ribosylation are involved in ubiquitination, it is tempting to suggest that Ube2q1/ubiquitin is ADP-ribosylated by Lem26 in such a way that canonical ubiquitination events are prevented. Further studies are needed to determine if Ube2q1 is a substrate of both MavL and Lem26.

In summary, this study provides crystallographic and binding data that point to a putative role for MavL in ADP-ribosylation. Structural comparisons of MavL with ADP-ribose binding macrodomains suggest reader or eraser activity against mono-ADP-ribosylated substrates. Although the biological role of MavL remains unclear, the studies presented here will guide further research toward this end.

Table 7. 1 Data collection and refinement statistics for His-MavL(42-435)

Data collection	
Space group	C121
Unit cell (Å)	97.6, 138.0, 108.5
Resolution (Å)	48.77–2.64 (2.66–2.64)
Total reflections	572585 (85399)
Unique reflections	82310 (13055)
Completeness (%)	99.30 (97.7)
Redundancy	6.96 (6.54)
Mean I/σ(I)	15.19 (1.37)
R_{merge} (%)	7.8 (138.7)
CC1/2	99.9 (64.9)
Refinement	
Resolution (Å)	45.14 – 2.64 (2.74 – 2.64)
No. of reflections	41712 (4052)
R_{work} / R_{free} (%)	21.7 / 25.5
No. of atoms / waters	8105 / 12
RMSD from ideal values	
Bond lengths (Å)	0.003
Bond angles (°)	0.62
Ramachandran plot (%)	
Favored	91.14
Allowed	8.02
Outlier	0.84
PDB ID	6OMI

8. Additional effectors investigated toward structure determination

8.1. LidL

8.1.1. Introduction

LidL is a paralogue of LpnE that contains 12 Sel1-like repeats (SLRs) and disrupts vacuolar trafficking. We aimed to structurally characterize LidL and compare its fold to that of LpnE, as this may highlight conserved and functionally relevant features of bacterial SLR proteins. Our interest in LidL was also inspired by a methods paper showing the crystallization of a fragment of this protein using a microfluidic device (Lee et al., 2014). The construct of LidL used in this publication included residues 27-192 out of 495 residues, corresponding to the first four SEl1-like repeats, and diffracted to 2.76 Å. However, the structure of this fragment has not been reported. Considering the nonspecific interactions mediated by the SLR super-helix, we felt that a structural understanding of the full-length protein would be more informative than a truncated fragment. That is, the architecture conferred by a full complement of SLRs will better describe the properties of LidL than a smaller set of residues.

8.1.2. Results

At the onset of this project, our database contained a LidL construct comprising residues 29-495. Removal of the N-terminal 28 residues eliminates a largely hydrophobic helix, which would likely interfere with the solubility and crystallization of this protein. Furthermore, analysis of LidL by the SignalP-5.0 server reveals a signal peptide at the N-terminus, with cleavage occurring between residues Ala25 and Leu26. Thus, LidL(29-495) was expressed with an N-terminal hexa-histidine tag from the pMCSG7 vector. Transformation of this plasmid into BL21(pLysS) led to an abundance of protein. Incubating the clarified lysate with Co^{2+} Talon beads for 1 hr at 4 °C proved to be an ineffective purification strategy, as nearly all recombinant protein was found in the flow-through (this was also true of the GST-tagged

LidL(29-495) expressed from pRL652 --this construct was not pursued further). His-LidL(29-495) could be purified, however, if the flow-through was added to NiNTA beads and incubated overnight at 4 °C. The yield from affinity chromatography was ~30 mg from 1 L, although some precipitation occurred immediately. Removal of precipitated protein by centrifugation reduced the final yield to ~24 mg. This protein was concentrated to 12 mg/ml and evaluated by size exclusion chromatography using a Biorad SEC650 column. Several peaks were observed on SEC650, with many contaminants eluting in the 12 – 17 ml range. LidL(29-495) eluted in a final peak centered at 17.75 ml. Intriguingly, SDS-PAGE shows these purified fractions to contain a doublet slightly above the 50 kDa marker (Figure 8.1). This may suggest that a small fragment of LidL(29-495) is cleaved during the purification.

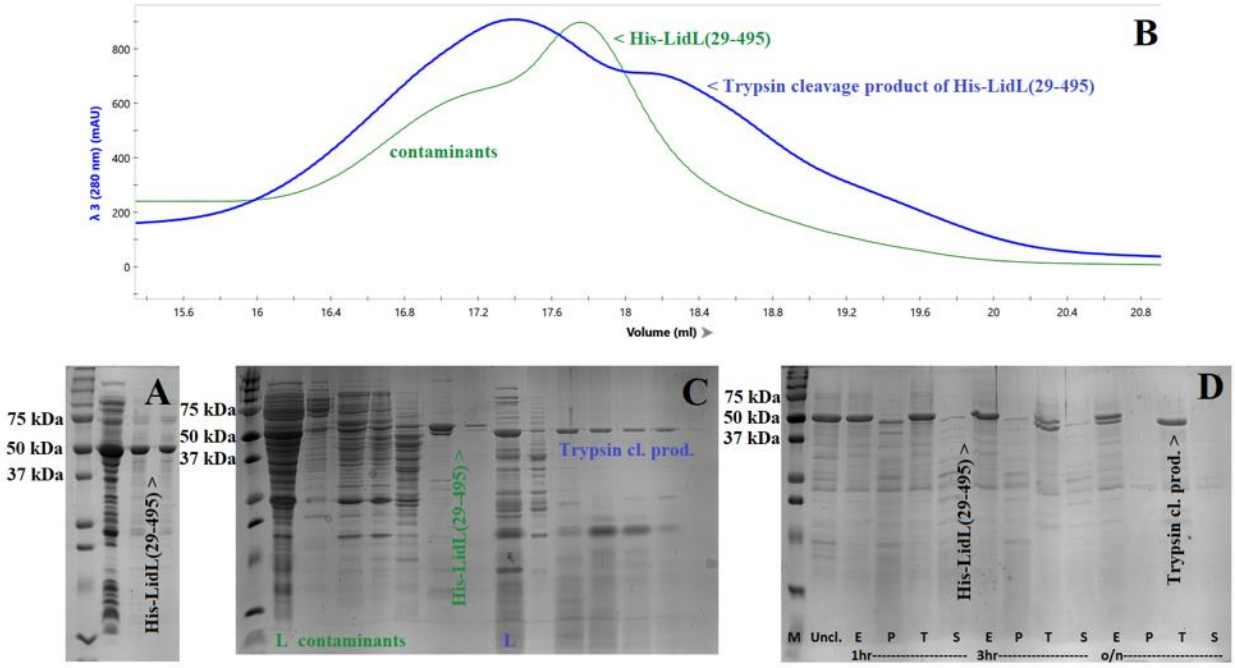

Figure 8. 1 Purification of His-LidL(29-495). (A) SDS-PAGE on affinity-purified His-LidL(29-495) after overnight incubation on NiNTA beads. From left to right, the lanes are: molecular weight marker, soluble protein, first elution and second elution. (B) Size exclusion chromatography on His-LidL(29-495) [Green] and the trypsin cleavage product [Blue] using a Biorad SEC650 column. (C) Fractions from the size exclusion runs depicted in B. The seven leftmost lanes correspond to undigested His-LidL(29-495) and the seven rightmost lanes correspond to the trypsin cleavage product. Note the His-LidL(29-495) doublet is no longer present in the digested protein. L = load. D) Limited proteolysis on His-LidL(29-495). M = marker, Uncl. = un-cleaved, E = elastase, P = proteinase K, T = trypsin, S = subtilisin. Incubation times are indicated below the enzyme abbreviations.

To explore the possibility of a stable proteolytic fragment for LidL(29-495), we carried out limited proteolysis on this construct using elastase, proteinase K, trypsin and subtilisin. Proteinase K and subtilisin degraded 95% of LidL(29-495) within the first three hours of digestion. Conversely, elastase and trypsin produced a stable cleavage product, with complete trypsin digestion occurring in 16 hrs. Importantly, the protein fragment obtained from trypsin and elastase digestion is similar in size to the smaller species observed in the LidL doublet (Figure 8.4B and C). It should be noted, however, that 15 % acrylamide gels were used in these experiments, making it difficult to distinguish proteins within the 47 – 55 kDa range. Thus, we cannot assume that the LidL fragment obtained from trypsin digestion is identical to that arising from unintended proteolysis. Still, a stable digestion fragment implies that a segment of LidL is accessible to trypsin and removal of this flexible region may promote crystallization.

Inspired by our results from limited proteolysis, we pursued structural studies on the LidL fragment obtained from trypsin digestion. Size exclusion chromatography on this fragment using a Biorad SEC650 column produced a major peak at 17.5 ml with a shoulder at 18.25 ml. SDS-PAGE analysis of protein containing fractions revealed that many of the contaminants had been eliminated during digestion. The LidL fragment appeared as a single band in both the major and shoulder peaks, indicating the presence of two oligomeric states or conformations for this protein.

LidL(29-495) was screened using CrystalScreen HT and pH Clear (Qiagen) at only 3 mg/ml because of its tendency to precipitate at higher concentrations. Trypsin-digested LidL was screened with CrystalScreen HT at 16 mg/ml. The lack of initial hits for undigested LidL(29-495) can be attributed to low concentration since nearly all conditions were clear (without precipitation) in these screens. Although trypsin-digested LidL could be brought to a higher concentration, the SDS-PAGE analysis of these purified fractions revealed the presence of contaminating proteins that may prevent crystallization (Figure 8.4C, Fractions 50 – 59).

Any number of explanations can be used to justify why a protein is recalcitrant to crystallization. For example, a protein-protein interaction interface may cause LidL to precipitate in the absence of its eukaryotic binding partner. Another issue could be the absence of a small molecule that stabilizes the protein in a specific conformation. One factor that warrants mentioning, however, is that only low concentrations (<12 mg/ml) of LidL were used

for crystallization. This was largely due to the tendency of this protein to precipitate, which may partly explain its low efficiency at binding the Ni^{2+} or Co^{2+} beads. Future efforts to address this issue of precipitation may include the use of alternative expression vectors and tags (eg. SUMO or MBP), an extensive buffer screening process or purification at low temperature (4 °C). In addition, the trypsin cleavage product of LidL may warrant further characterization by mass spectrometry. This would allow the residues representing this fragment to be identified and cloned into an expression vector. Production of this recombinant fragment may facilitate crystallization if undigested protein otherwise poisons the contact interfaces.

8.1.3. Discussion

LidL is a Sel1-like repeat (SLR) protein that allows *L. pneumophila* to influence vacuolar trafficking in an infected host cell. Despite being unable to crystallize LidL(29-495), our studies on this protein provide some insight into purification strategies. Of note is the stable cleavage product we obtained from trypsin digestion. Only a small peptide fragment is cleaved by trypsin to render a band that migrates just below the 50 kDa marker on SDS-PAGE. Trypsin is known to cleave peptides on the C-terminal side of lysine and arginine residues. Secondary structure analysis (PsiPred) of LidL shows several potential cleavage sites within loop regions that would give rise to a product in the 46 – 48 kDa range. These include Lys67, Lys425 and Arg426, which produce LidL(68-495), His-LidL(29-425) and His-LidL(29-426), respectively. Since no western blot or mass spectrometry was carried out for the LidL trypsin cleavage product, the residue composition of this species remains unclear.

Digestion of LidL with trypsin serves the dual purpose of eliminating (1) the doublet observed on SDS-PAGE for undigested His-LidL(29-495) and (2) high molecular weight contaminants. Removal of the doublet suggests that trypsin cleaves at, or slightly beyond, the original point of separation; if the original cleavage site is retained in the trypsin digested fragment, we should expect this product to be partially cleaved as well. It is also possible that trypsin degrades or inactivates the enzyme that originally cleaves His-LidL(29-495). This would abolish any relationship between the site of endogenous- and trypsin-cleavage. Importantly, our construct of LidL starts beyond the predicted signal peptide. The doublet

observed in the purification of LidL(29-495) is therefore unlikely to represent the cleavage of a signal peptide. The ability of trypsin to produce a single LidL species bodes well for crystallization, as homogeneity of the sample is essential for this process. Removal of the high molecular weight contaminants by trypsin is evidenced by size exclusion chromatography. The broad peak centered at 13.75 ml disappears completely in the chromatogram of trypsin digested LidL. SDS-PAGE analysis of these fractions confirms the removal of contaminants. Identifying the residue composition of the LidL trypsin cleavage product may lead to the rational design of new constructs that resist proteolysis and are amenable to crystallization.

8.2. LirF

8.2.1. Introduction

LirF is an *L. pneumophila* effector protein of unknown function. We sought to determine the structure of this protein to learn about its function. Analysis of LirF using the PsiPred secondary structure prediction tool shows a combination of α-helices and β-sheets. Two constructs of LirF were explored at the onset of this project: LirF(2-521) and LirF(2-331). The C-terminal half of LirF contains several extended loops interspersed with short secondary structure elements (SSEs), while the N-terminal half is dominated by SSEs. We hypothesized that a construct of LirF lacking the C-terminal half may be more stable and therefore amenable to crystallization.

8.2.2. Results

Initial efforts to purify LirF(2-521) and LirF(2-331) were carried out using the pMCSG7 vector, which confers an N-terminal hexa-histidine tag. Both constructs expressed nicely in BL21(DE3)pLysS, but only the full-length protein (LirF(2-521)) was soluble and could be purified on Ni^{2+} and Co^{2+} beads (Figure 8.1). The fact that LirF(2-331) was insoluble may suggest that removal of the C-terminal region exposes a hydrophobic surface giving rise to precipitation. Alternatively, the removed segment may support the structural integrity of LirF. Whatever the reason, the full-length construct (LirF(2-521)) showed promise for downstream structural studies and became the focus of future efforts toward this end.

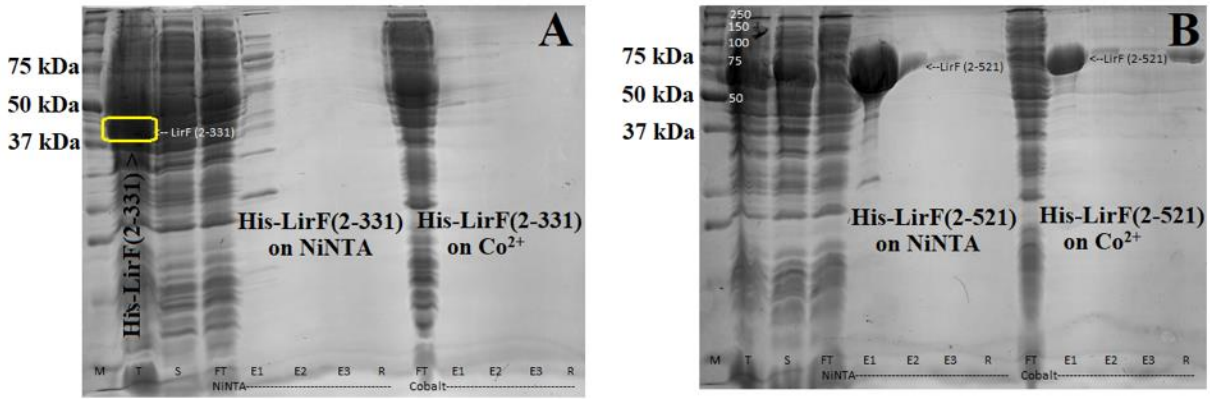

Figure 8. 2 Purification of His-LirF(2-331) and His-LirF(2-521) on Ni^{2+} and Co^{2+}. (A) SDS-PAGE monitoring of the His-LirF(2-331) purification. (B) SDS-PAGE monitoring of the His-LirF(2-521) purification. Note that His-LirF(2-331) is only present in the total lysate (yellow box) and not in the soluble fraction. His-LirF(2-521) is soluble and purifies on both NiNTA and Co^{2+}.

Size exclusion chromatography on His-LirF(2-521) was carried out using a GE Superdex75 column. (Efforts to cleave the N-terminal hexa-histidine tag from LirF(2-521) were met with little success. Briefly, tobacco etch virus (TEV) protease was added to LirF samples at a 1:50 ratio and incubated overnight at 4 °C. This process was repeated for three consecutive days with no noticeable cleavage. These findings may suggest the TEV cleavage site is occluded in properly folded LirF.) Consequently, tagged protein was concentrated to 10 mg/ml using a Millipore 30 kDa molecular weight cut-off (MWCO) centrifugal filter and a volume of 500 µl was injected onto the column. The protein eluted in a major peak at ~17.5 ml, with a minor peak preceding it that also contains some LirF and contaminating protein. Major peak fractions were collected and concentrated to 11 mg/ml and screened using pH Clear, MbClassI and the ClassicsII 96-well suites. Excess precipitation was observed in many conditions with no initial crystals to optimize.

Although LirF(2-521) could be purified on affinity and size exclusion chromatography, it also had a tendency to precipitate during dialysis and in solution. We set up a screen at 4 °C to mitigate this tendency during crystal growth but observed no crystals. Increasing the final salt concentration to 500 mM NaCl reduced the precipitation that occurs in a concentrated LirF stock overnight at 4 °C. We tried to circumvent the issue of LirF precipitation using limited proteolysis. Our hope was to obtain a stable proteolytic fragment of LirF that would resist

precipitation and crystallize more readily. Elastase, proteinase K, trypsin and subtilisin were used to cleave LirF in an enzyme:substrate ratio of 1:1000. Trypsin was found to produce a stable ~42 kDa cleavage product of LirF(2-521). Elastase had no noticeable effect on LirF(2-521), whereas proteinase K and subtilisin completely degraded the protein (Figure 8.2A). The stable product obtained from trypsin digestion was characterized by western blot using an anti-His-HRP conjugated antibody (Figure 8.2B). Since the ~42 kDa band produced a signal using this antibody, we concluded that our N-terminal hexa-histidine tag was intact in the proteolyzed fragment. This implied that cleavage was occurring at the C-terminus of the protein. The most likely site of trypsin cleavage is Lys336, which would eliminate a peptide of 18.0 kDa from LirF(2-521), rendering an N-terminal fragment of 41.4 kDa.

We then purified the trypsin cleavage product of LirF(2-521) by size exclusion chromatography (GE Superdex200 column). The trypsin digested LirF(2-521) eluted in a peak centered at 16.75 ml, which is slightly downshifted relative to the undigested protein (Figure 8.3A, green trace; C, peak fractions). Despite extensive efforts to crystallize both full-length and trypsin digested LirF, no crystal hits were obtained. This project was not pursued further at this stage, but future efforts toward the structural characterization of LirF should consider the utility of a construct terminating at Arg361. Complete digestion of LirF(2-521) by trypsin proved difficult, and having a construct corresponding to this digested fragment may aid in crystallization. In addition, the six N-terminal residues of LirF are predicted to represent a re-entrant helix, with residue eight (Thr8) being disordered altogether. Exploring N-terminal truncations may therefore reveal a LirF construct that resists precipitation and crystallizes readily.

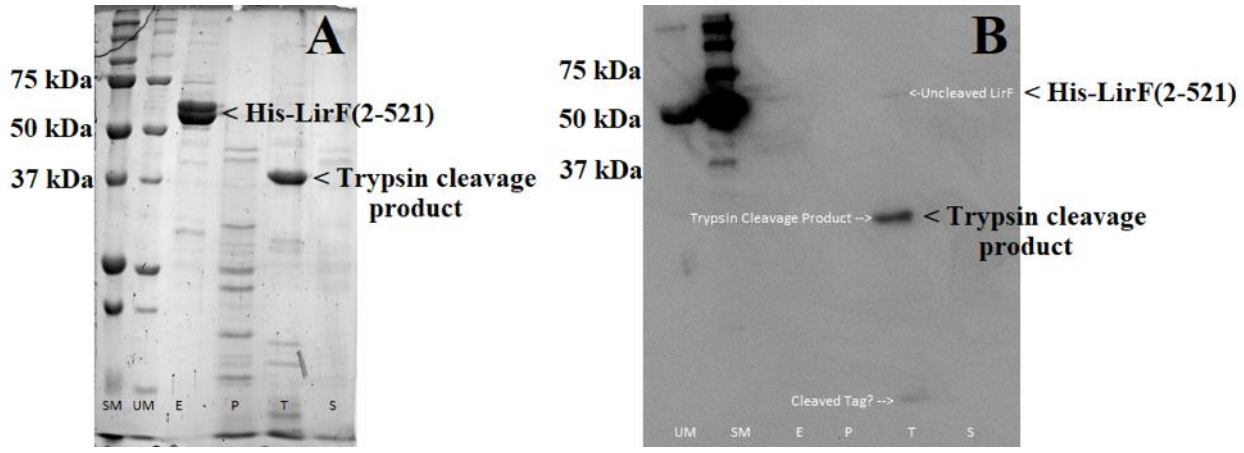

Figure 8. 3 Limited proteolysis on His-LirF(2-521) reveals a stable trypsin cleavage product.(A) SDS-PAGE monitoring of limited proteolysis after 24 hrs at 4 °C. SM = stained molecular weight marker, UM = unstained molecular weight marker, E = elastase, P = proteinase K, T = trypsin, S = subtilisin. (B) Western blot on LirF(2-521) after being digested with various enzymes for 24 hrs at 4 °C.

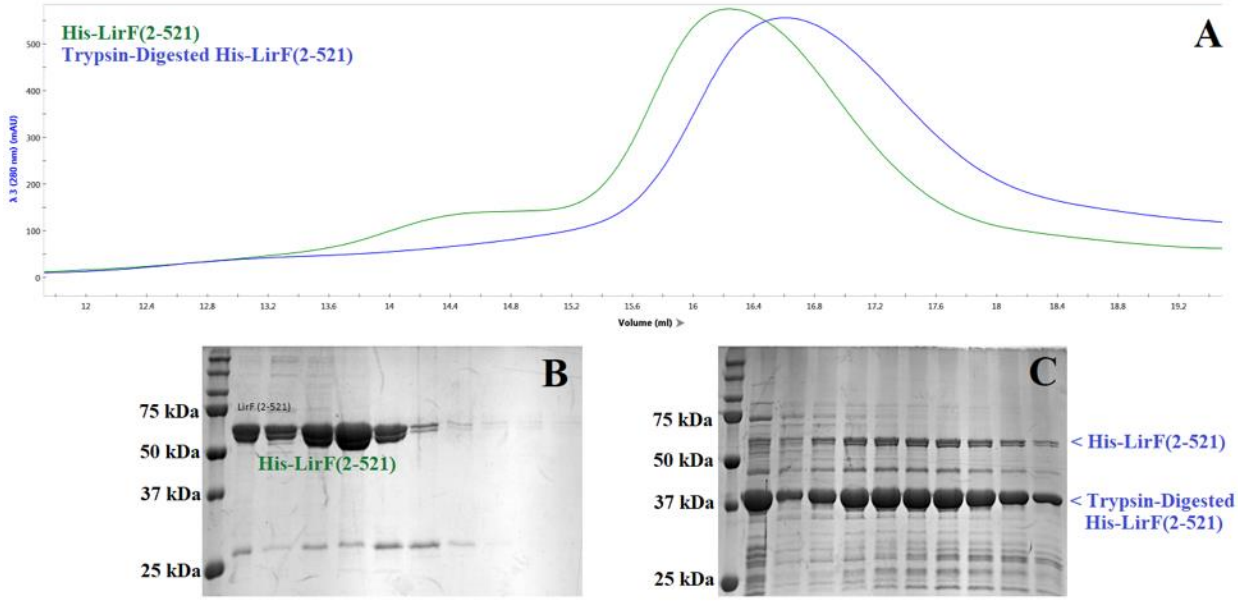

Figure 8. 4 Size exclusion chromatography of His-LirF(2-521) and trypsin digested His-LirF(2-521) on a GE Superdex200 column.(A) Overlay of chromatograms obtained for trypsin-digested [Blue] and undigested [Green] His-LirF(2-521). (B) SDS-PAGE analysis of peak fraction from undigested His-LirF(2-521). (C) SDS-PAGE analysis of peak fractions from trypsin-digested LirF(2-521).

8.2.3. LirF Discussion

LirF is an *L. pneumophila* effector protein of unknown function. Although we were unable to crystallize and solve the structure of this protein, we did make some discoveries to aid in future purification strategies. For example, our C-terminal truncation, LirF(2-331), was completely insoluble, while the full-length protein, LirF(2-521) was soluble and easily purified on Ni^{2+} and Co^{2+}. This could indicate that removal of the LirF C-terminus exposes a hydrophobic interface giving rise to precipitation. Alternatively, proper folding of LirF may depend on the presence of its C-terminal half. Secondary structure analysis reveals an extended loop region from Asp327 to Cys376 with only a few short helices and strands predicted within it. This loop may connect folded N- and C-terminal domains of LirF that are represented by a greater length and number of secondary structure elements. Systematically exploring C-terminal truncations of full-length LirF may help us to understand how solubility is promoted by C-terminal extensions beyond residue 331. In addition, the eight N-terminal and 20 C-terminal residues of LirF are predicted to be somewhat disordered. It is therefore possible that removal of these residues would facilitate the crystallization of LirF.

We found that digestion of His-LirF(2-521) with trypsin produces a stable ~42 kDa cleavage product that retains the hexa-histidine tag. The most likely cleavage point is Lys336, as this residue is not within a secondary structure element or flanked by a C-terminal proline. Inspection of the disorder plot generated for LirF shows that disordered residues exist in the 125 – 150 and 450 – 521 regions. On this basis, the trypsin-digested fragment may be more likely to crystallize than its full-length counterpart. This limited proteolysis study can also be used to infer how the C-terminus affects LirF stability, because the trypsin digestion occurs at the C-terminus of LirF and does not impede its solubility. If trypsin does indeed cleave at Lys336, the short strand between Asn331 and Lys336 may be instrumental in facilitating LirF solubility. The intervening sequence is comprised of residues LKFKK, which may promote solubility via the charged lysine sidechains. It should also be noted that Lys387 is a potential point of trypsin cleavage, although this residue would give rise to a 47 kDa product, which seems a little large based on the SDS-PAGE results. Still, cleavage at this position implicates a greater number of residues in promoting solubility. Only one arginine and lysine residue exist within this intervening region and both are flanked by a proline residue at their C-terminus.

This is convenient, as it limits our choices for the trypsin cleavage site to Lys336 and Lys387. Designing constructs of LirF that terminate at Lys336 and Lys387 may verify the speculations presented here regarding the C-terminal affect on solubility.

9. Conclusions and Future Directions

Legionella pneumophila is an intracellular pathogen that is increasing in virulence by developing its effector repertoire. This bacterium uses a large arsenal of secreted effector proteins to promote replication inside a wide range of protozoan species. Many of these effectors are acquired from eukaryotes by horizontal gene transfer and interfere with host processes by molecular mimicry. Functional redundancy of *L. pneumophila* effectors makes structure determination an attractive avenue of investigation into these proteins. In this dissertation, crystal structures are presented for three *L. pneumophila* effectors that were functionally uncharacterized prior to this work: LpnE, MavE and MavL.

We have found that LpnE relies on its N-terminal signal peptide to promote *cis*-Golgi localization in HEK293 cells. This same sequence likely anchors LpnE to the LCV during infection. Previous studies have shown LpnE to recruit OCRL to the LCV and restrict the intracellular replication of *L. pneumophila*. Our LpnE-OCRL binding studies reveal that this interaction occurs between a polyproline motif on OCRL and a central helix (3-α_B) of LpnE. These findings satisfy our primary objective to explain the binding of LpnE with OCRL, although further studies are still needed to map this interaction in molecular detail. For example, co-crystallization of LpnE with OCRL or the polyproline peptide fragment would reveal the residues involved in this interaction. Such studies expand our appreciation of the strategies employed by SLR proteins to engage in protein-protein interactions and may provide clues to understanding how replication is inhibited by OCRL recruitment to the LCV. Considering the promiscuity of SLR folds, it is likely that LpnE uses unique binding partners to facilitate host cell entry. Screening for LpnE targets on macrophage cell membranes may provide insight into the process of invasion.

The crystal structure of MavE(39-172) shares a twinned two-helix bundle core domain with the translation initiation factor eIF2Bα, methylthioribose-1-phosphate isomerase, ribose-1,5-bisphosphate isomerase and grass pollen allergen Phlp-5b. A fifth helix (B) on MavE is absent in all four of these homologous proteins, suggesting a distinguishing functional role for this structural feature. Intriguingly, a surface exposed loop connecting helix B to the two-helix bundle harbors an NPxY motif. Considering the ability of NPxY motifs to bind adaptor proteins, these structural observations provide a convenient hypothesis for the role MavE plays

in vacuolar trafficking. For example, MavE may bind tethering factors that direct the LCV away from lysosomes. These observations satisfy our objective to describe structural features in MavE that are significant to its function. Additional efforts are required to identify the molecular targets of MavE and verify the role of this NPxY motif in such interactions.

Our MavL(42-435) structure has a similar fold to that of ADP-ribose binding macrodomains and harbors a pair of aspartate residues in the correct position to facilitate glycohydrolase activity. These findings satisfy our objective to arrive at a basic functional interpretation of MavL using its molecular structure. Isothermal titration calorimetry (ITC) verified the binding of ADP-ribose to MavL(42-435), revealing a dissociation constant of 13 μM. Although these findings offer a convenient explanation for the inhibition of Lem26 by MavL, it remains unclear whether MavL indeed possesses glycohydrolase activity. A better understanding of Lem26 functionality may explain the link between ADP-ribosylation, MavL and possibly its ability to bind the E2 ubiquitin conjugating enzyme, Ube2q1.

References

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Echols, N., Headd, J.J., Hung, L.-W., Jain, S., Kapral, G.J., Grosse Kunstleve, R.W., et al. (2011). The Phenix software for automated determination of macromolecular structures. Methods *55*, 94–106.

Agrotis, A., Pengo, N., Burden, J.J., and Ketteler, R. (2019). Redundancy of human ATG4 protease isoforms in autophagy and LC3/GABARAP processing revealed in cells. Autophagy.

Akturk, A., Wasilko, D.J., Wu, X., Liu, Y., Zhang, Y., Qiu, J., Luo, Z.-Q., Reiter, K.H., Brzovic, P.S., Klevit, R.E., et al. (2018). Mechanism of phosphoribosyl-ubiquitination mediated by a single Legionella effector. Nature *557*, 729–733.

Allan, B.B., Moyer, B.D., and Balch, W.E. (2000a). Rab1 recruitment of p115 into a cis-SNARE complex: programming budding COPII vesicles for fusion. Science *289*, 444–448.

Allan, B.B., Moyer, B.D., and Balch, W.E. (2000b). Rab1 recruitment of p115 into a cis-SNARE complex: programming budding COPII vesicles for fusion. Science *289*, 444–448.

Allgood, S.C., and Neunuebel, M.R. (2018). The recycling endosome and bacterial pathogens. Cell. Microbiol. *20*, e12857.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Arasaki, K., Toomre, D.K., and Roy, C.R. (2012). The Legionella pneumophila effector DrrA is sufficient to stimulate SNARE-dependent membrane fusion. Cell Host Microbe *11*, 46–57.

Arasaki, K., Kimura, H., Tagaya, M., and Roy, C.R. (2018). Legionella remodels the plasma membrane-derived vacuole by utilizing exocyst components as tethers. J. Cell Biol. *217*, 3863–3872.

Bailey, L.J., Sheehy, K.M., Dominik, P.K., Liang, W.G., Rui, H., Clark, M., Jaskolowski, M., Kim, Y., Deneka, D., Tang, W.-J., et al. (2018). Locking the Elbow: Improved Antibody Fab Fragments as Chaperones for Structure Determination. J. Mol. Biol. *430*, 337–347.

Bandyopadhyay, P., Sumer, E.U., Jayakumar, D., Liu, S., Xiao, H., and Steinman, H.M. (2012). Implication of proteins containing tetratricopeptide repeats in conditional virulence phenotypes of Legionella pneumophila. J. Bacteriol. *194*, 3579–3588.

Banfield, D.K. (2011). Mechanisms of protein retention in the Golgi. Cold Spring Harb Perspect Biol *3*, a005264.

Bärlocher, K., Welin, A., and Hilbi, H. (2017). Formation of the Legionella Replicative Compartment at the Crossroads of Retrograde Trafficking. Front Cell Infect Microbiol *7*, 482.

Benabdi, S., Peurois, F., Nawrotek, A., Chikireddy, J., Cañeque, T., Yamori, T., Shiina, I., Ohashi, Y., Dan, S., Rodriguez, R., et al. (2017). Family-wide Analysis of the Inhibition of Arf Guanine Nucleotide Exchange Factors with Small Molecules: Evidence of Unique Inhibitory Profiles. Biochemistry *56*, 5125–5133.

Benin, A.L., Benson, R.F., and Besser, R.E. (2002). Trends in legionnaires disease, 1980-1998: declining mortality and new patterns of diagnosis. Clin. Infect. Dis. *35*, 1039–1046.

Benson, R.F., and Fields, B.S. (1998). Classification of the genus Legionella. Semin Respir Infect *13*, 90–99.

Beyrakhova, K., Li, L., Xu, C., Gagarinova, A., and Cygler, M. (2018). Legionella pneumophila effector Lem4 is a membrane-associated protein tyrosine phosphatase. J. Biol. Chem. *293*, 13044–13058.

Bragg, W.L. (1912). *The Specular Reflection of X-rays.* Nature *90*, 410.

Bragg, W.L. (1913). The diffraction of short electromagnetic waves by a crystal. Mathematical Proceedings of the Cambridge Philosophical Society *17*, 43–57.

Campodonico, E.M., Roy, C.R., and Ninio, S. (2016). Legionella pneumophila Type IV Effectors YlfA and YlfB Are SNARE-Like Proteins that Form Homo- and Heteromeric Complexes and Enhance the Efficiency of Vacuole Remodeling. PLoS ONE *11*, e0159698.

de Carvalho, T.M.U., Barrias, E.S., and de Souza, W. (2015). Macropinocytosis: a pathway to protozoan infection. Front Physiol *6*, 106.

Cerveny, L., Straskova, A., Dankova, V., Hartlova, A., Ceckova, M., Staud, F., and Stulik, J. (2013). Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. Infect. Immun. *81*, 629–635.

Chen, W.J., Goldstein, J.L., and Brown, M.S. (1990). NPXY, a sequence often found in cytoplasmic tails, is required for coated pit-mediated internalization of the low density lipoprotein receptor. J. Biol. Chem. *265*, 3116–3123.

Chen, Y., Tascón, I., Neunuebel, M.R., Pallara, C., Brady, J., Kinch, L.N., Fernández-Recio, J., Rojas, A.L., Machner, M.P., and Hierro, A. (2013). Structural basis for Rab1 de-AMPylation by the Legionella pneumophila effector SidD. PLoS Pathog. *9*, e1003382.

Chernov, A.A., and Komatsu, H. (1995). Principles of Crystal Growth in Protein Crystallization. In Science and Technology of Crystal Growth, J.P. van der Eerden, and O.S.L. Bruinsma, eds. (Dordrecht: Springer Netherlands), pp. 329–353.

Chetrit, D., Hu, B., Christie, P.J., Roy, C.R., and Liu, J. (2018). A unique cytoplasmic ATPase complex defines the Legionella pneumophila type IV secretion channel. Nat Microbiol *3*, 678–686.

Christoforidis, S., McBride, H.M., Burgoyne, R.D., and Zerial, M. (1999). The Rab5 effector EEA1 is a core component of endosome docking. Nature *397*, 621–625.

Cigler, M., Müller, T.G., Horn-Ghetko, D., von Wrisberg, M.-K., Fottner, M., Goody, R.S., Itzen, A., Müller, M.P., and Lang, K. (2017). Proximity-Triggered Covalent Stabilization of Low-Affinity Protein Complexes In Vitro and In Vivo. Angew. Chem. Int. Ed. Engl. *56*, 15737–15741.

Cirillo, S.L., Lum, J., and Cirillo, J.D. (2000). Identification of novel loci involved in entry by Legionella pneumophila. Microbiology (Reading, Engl.) *146 (Pt 6)*, 1345–1359.

Cliff, M.J., Harris, R., Barford, D., Ladbury, J.E., and Williams, M.A. (2006). Conformational diversity in the TPR domain-mediated interaction of protein phosphatase 5 with Hsp90. Structure *14*, 415–426.

Coudroy, G., Gburek, J., Kozyraki, R., Madsen, M., Trugnan, G., Moestrup, S.K., Verroust, P.J., and Maurice, M. (2005). Contribution of cubilin and amnionless to processing and membrane targeting of cubilin-amnionless complex. J. Am. Soc. Nephrol. *16*, 2330–2337.

Das, A.K., Cohen, P.W., and Barford, D. (1998). The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. EMBO J. *17*, 1192–1199.

De Matteis, M.A., Staiano, L., Emma, F., and Devuyst, O. (2017). The 5-phosphatase OCRL in Lowe syndrome and Dent disease 2. Nat Rev Nephrol *13*, 455–470.

DebRoy, S., Dao, J., Söderberg, M., Rossier, O., and Cianciotto, N.P. (2006). Legionella pneumophila type II secretome reveals unique exoproteins and a chitinase that promotes bacterial persistence in the lung. Proc. Natl. Acad. Sci. U.S.A. *103*, 19146–19151.

Derewenda, Z.S., and Godzik, A. (2017). The "Sticky Patch" Model of Crystallization and Modification of Proteins for Enhanced Crystallizability. Methods Mol. Biol. *1607*, 77–115.

Discola, K.F., Förster, A., Boulay, F., Simorre, J.-P., Attree, I., Dessen, A., and Job, V. (2014). Membrane and chaperone recognition by the major translocator protein PopB of the type III secretion system of Pseudomonas aeruginosa. J. Biol. Chem. *289*, 3591–3601.

Dong, N., Niu, M., Hu, L., Yao, Q., Zhou, R., and Shao, F. (2016). Modulation of membrane phosphoinositide dynamics by the phosphatidylinositide 4-kinase activity of the Legionella LepB effector. Nat Microbiol *2*, 16236.

Dong, X., Shen, D., Wang, X., Dawson, T., Li, X., Zhang, Q., Cheng, X., Zhang, Y., Weisman, L.S., Delling, M., et al. (2010). PI(3,5)P(2) controls membrane trafficking by direct activation of mucolipin Ca(2+) release channels in the endolysosome. Nat Commun *1*, 38.

Dong, Y., Mu, Y., Xie, Y., Zhang, Y., Han, Y., Zhou, Y., Wang, W., Liu, Z., Wu, M., Wang, H., et al. (2018). Structural basis of ubiquitin modification by the Legionella effector SdeA. Nature *557*, 674–678.

Dressman, M.A., Olivos-Glander, I.M., Nussbaum, R.L., and Suchy, S.F. (2000). Ocrl1, a PtdIns(4,5)P(2) 5-phosphatase, is localized to the trans-Golgi network of fibroblasts and epithelial cells. J. Histochem. Cytochem. *48*, 179–190.

Elwell, C., and Engel, J. (2018). Emerging Role of Retromer in Modulating Pathogen Growth. Trends Microbiol. *26*, 769–780.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Ensminger, A.W. (2016). Legionella pneumophila, armed to the hilt: justifying the largest arsenal of effectors in the bacterial world. Curr. Opin. Microbiol. *29*, 74–80.

Eschenfeldt, W.H., Lucy, S., Millard, C.S., Joachimiak, A., and Mark, I.D. (2009). A family of LIC vectors for high-throughput cloning and purification of proteins. Methods Mol. Biol. *498*, 105–115.

Evans, P., and McCoy, A. (2008). An introduction to molecular replacement. Acta Crystallogr D Biol Crystallogr *64*, 1–10.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr. D Biol. Crystallogr. *69*, 1204–1214.

Ewald, P.P. (1969). Introduction to the dynamical theory of X-ray diffraction. Acta Cryst A *25*, 103–108.

Flayhan, A., Bergé, C., Baïlo, N., Doublet, P., Bayliss, R., and Terradot, L. (2015). The structure of Legionella pneumophila LegK4 type four secretion system (T4SS) effector reveals a novel dimeric eukaryotic-like kinase. Sci Rep *5*, 14602.

Fookes, M., Schroeder, G.N., Langridge, G.C., Blondel, C.J., Mammina, C., Connor, T.R., Seth-Smith, H., Vernikos, G.S., Robinson, K.S., Sanders, M., et al. (2011). Salmonella bongori provides insights into the evolution of the Salmonellae. PLoS Pathog. *7*, e1002191.

Freyer, M.W., and Lewis, E.A. (2008). Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. Methods Cell Biol. *84*, 79–113.

Friedrich, W., and Knipping, P. (1913). Interference Phenomena with rontgen rays. Annals of Physics *41*, 971–988.

Garcia-Vallve, S., Guzman, E., Montero, M.A., and Romeu, A. (2003). HGT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. Nucleic Acids Res. *31*, 187–189.

Garman, E.F. (2014). Developments in x-ray crystallographic structure determination of biological macromolecules. Science *343*, 1102–1108.

Ghosal, D., Chang, Y.-W., Jeong, K.C., Vogel, J.P., and Jensen, G.J. (2017). In situ structure of the Legionella Dot/Icm type IV secretion system by electron cryotomography. EMBO Rep. *18*, 726–732.

Gomez-Valero, L., Rusniok, C., Cazalet, C., and Buchrieser, C. (2011). Comparative and functional genomics of legionella identified eukaryotic like proteins as key players in host-pathogen interactions. Front Microbiol *2*, 208.

Gomez-Valero, L., Rusniok, C., Carson, D., Mondino, S., Pérez-Cobas, A.E., Rolando, M., Pasricha, S., Reuter, S., Demirtas, J., Crumbach, J., et al. (2019). More than 18,000 effectors in the Legionella genus genome provide multiple, independent combinations for replication in human cells. Proc. Natl. Acad. Sci. U.S.A.

Gong, X., Zhao, X., Zhang, W., Wang, J., Chen, X., Hameed, M.F., Zhang, N., and Ge, H. (2018). Structural characterization of the hypothetical protein Lpg2622, a new member of the C1 family peptidases from Legionella pneumophila. FEBS Lett. *592*, 2798–2810.

Goody, R.S., Müller, M.P., and Wu, Y.-W. (2017). Mechanisms of action of Rab proteins, key regulators of intracellular vesicular transport. Biol. Chem. *398*, 565–575.

Grishin, A.M., Beyrakhova, K.A., and Cygler, M. (2015). Structural insight into effector proteins of Gram-negative bacterial pathogens that modulate the phosphoproteome of their host. Protein Sci. *24*, 604–620.

Grochulski, P., Fodje, M.N., Gorin, J., Labiuk, S.L., and Berg, R. (2011). Beamline 08ID-1, the prime beamline of the Canadian Macromolecular Crystallography Facility. J Synchrotron Radiat *18*, 681–684.

Grohmann, E., Keller, W., and Muth, G. (2017). Mechanisms of Conjugative Transfer and Type IV Secretion-Mediated Effector Transport in Gram-Positive Bacteria. Curr. Top. Microbiol. Immunol. *413*, 115–141.

Han, S., Craig, J.A., Putnam, C.D., Carozzi, N.B., and Tainer, J.A. (1999). Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex. Nat. Struct. Biol. *6*, 932–936.

Hardiman, C.A., and Roy, C.R. (2014). AMPylation is critical for Rab1 localization to vacuoles containing Legionella pneumophila. MBio *5*, e01035-01013.

Harrison, M.A., and Muench, S.P. (2018). The Vacuolar ATPase - A Nano-scale Motor That Drives Cell Biology. Subcell. Biochem. *87*, 409–459.

Hemrajani, C., Berger, C.N., Robinson, K.S., Marchès, O., Mousnier, A., and Frankel, G. (2010). NleH effectors interact with Bax inhibitor-1 to block apoptosis during enteropathogenic Escherichia coli infection. Proc. Natl. Acad. Sci. U.S.A. *107*, 3129–3134.

Hendrickson, W.A. (2014). Anomalous Diffraction in Crystallographic Phase Evaluation. Q Rev Biophys *47*, 49–93.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J. *9*, 1665–1672.

Herak-Kramberger, C.M., Brown, D., and Sabolić, I. (1998). Cadmium inhibits vacuolar H(+)-ATPase and endocytosis in rat kidney cortex. Kidney Int. *53*, 1713–1726.

Hervet, E., Charpentier, X., Vianney, A., Lazzaroni, J.-C., Gilbert, C., Atlan, D., and Doublet, P. (2011). Protein kinase LegK2 is a type IV secretion system effector involved in endoplasmic reticulum recruitment and intracellular replication of Legionella pneumophila. Infect. Immun. *79*, 1936–1950.

Hoffmann, C., Finsel, I., Otto, A., Pfaffinger, G., Rothmeier, E., Hecker, M., Becher, D., and Hilbi, H. (2014). Functional analysis of novel Rab GTPases identified in the proteome of purified Legionella-containing vacuoles from macrophages. Cell. Microbiol. *16*, 1034–1052.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545-549.

Hope, H. (1988). Cryocrystallography of biological macromolecules: a generally applicable method. Acta Crystallogr., B *44 (Pt 1)*, 22–26.

Hsu, F., Zhu, W., Brennan, L., Tao, L., Luo, Z.-Q., and Mao, Y. (2012). Structural basis for substrate recognition by a unique Legionella phosphoinositide phosphatase. Proc. Natl. Acad. Sci. U.S.A. *109*, 13567–13572.

Hsu, F., Luo, X., Qiu, J., Teng, Y.-B., Jin, J., Smolka, M.B., Luo, Z.-Q., and Mao, Y. (2014). The Legionella effector SidC defines a unique family of ubiquitin ligases important for bacterial phagosomal remodeling. Proc. Natl. Acad. Sci. U.S.A. *111*, 10538–10543.

Huang, L., Boyd, D., Amyot, W.M., Hempstead, A.D., Luo, Z.-Q., O'Connor, T.J., Chen, C., Machner, M., Montminy, T., and Isberg, R.R. (2011). The E Block motif is associated with Legionella pneumophila translocated substrates. Cell. Microbiol. *13*, 227–245.

Jeong, H., Sim, H.J., Song, E.K., Lee, H., Ha, S.C., Jun, Y., Park, T.J., and Lee, C. (2016). Crystal structure of SEL1L: Insight into the roles of SLR motifs in ERAD pathway. Sci Rep *6*, 20261.

Jia, Q., Lin, Y., Gou, X., He, L., Shen, D., Chen, D., Xie, W., and Lu, Y. (2018). Legionella pneumophila effector WipA, a bacterial PPP protein phosphatase with PTP activity. Acta Biochim. Biophys. Sin. (Shanghai) *50*, 547–554.

Job, V., Matteï, P.-J., Lemaire, D., Attree, I., and Dessen, A. (2010). Structural basis of chaperone recognition of type III secretion system minor translocator proteins. J. Biol. Chem. *285*, 23224–23232.

Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132.

Kagan, J.C., and Roy, C.R. (2002). Legionella phagosomes intercept vesicular traffic from endoplasmic reticulum exit sites. Nat. Cell Biol. *4*, 945–954.

Kalayil, S., Bhogaraju, S., Bonn, F., Shin, D., Liu, Y., Gan, N., Basquin, J., Grumati, P., Luo, Z.-Q., and Dikic, I. (2018). Insights into catalysis and function of phosphoribosyl-linked serine ubiquitination. Nature *557*, 734–738.

Kaneko, T., Stogios, P.J., Ruan, X., Voss, C., Evdokimova, E., Skarina, T., Chung, A., Liu, X., Li, L., Savchenko, A., et al. (2018). Identification and characterization of a large family of superbinding bacterial SH2 domains. Nat Commun *9*, 4549.

Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. Science *336*, 1030–1033.

Kashiwagi, K., Takahashi, M., Nishimoto, M., Hiyama, T.B., Higo, T., Umehara, T., Sakamoto, K., Ito, T., and Yokoyama, S. (2016). Crystal structure of eukaryotic translation initiation factor 2B. Nature *531*, 122–125.

Kim, D.W., Lenzen, G., Page, A.-L., Legrain, P., Sansonetti, P.J., and Parsot, C. (2005). The Shigella flexneri effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. Proc. Natl. Acad. Sci. U.S.A. *102*, 14046–14051.

Krojer, T., Pike, A.C.W., and von Delft, F. (2013). Squeezing the most from every crystal: the fine details of data collection. Acta Crystallogr. D Biol. Crystallogr. *69*, 1303–1313.

Ku, B., Lee, K.-H., Park, W.S., Yang, C.-S., Ge, J., Lee, S.-G., Cha, S.-S., Shao, F., Heo, W.D., Jung, J.U., et al. (2012). VipD of Legionella pneumophila targets activated Rab5 and Rab22 to interfere with endosomal trafficking in macrophages. PLoS Pathog. *8*, e1003082.

Kubori, T., Kitao, T., Ando, H., and Nagai, H. (2018). LotA, a Legionella deubiquitinase, has dual catalytic activity and contributes to intracellular growth. Cell. Microbiol. *20*, e12840.

Kwak, M.-J., Kim, J.D., Kim, H., Kim, C., Bowman, J.W., Kim, S., Joo, K., Lee, J., Jin, K.S., Kim, Y.-G., et al. (2017). Architecture of the type IV coupling protein complex of Legionella pneumophila. Nat Microbiol *2*, 17114.

Kwon, D.H., and Song, H.K. (2018). A Structural View of Xenophagy, a Battle between Host and Microbes. Mol. Cells *41*, 27–34.

Larsen, F.K. (1995). Diffraction studies of crystals at low temperatures – crystallography below 77 K. Acta Cryst B *51*, 468–482.

Lee, P.-C., and Machner, M.P. (2018). The Legionella Effector Kinase LegK7 Hijacks the Host Hippo Pathway to Promote Infection. Cell Host Microbe *24*, 429-438.e6.

Lee, M.J.Y., Faucher, F., and Jia, Z. (2014). Growth of Diffraction-Quality Protein Crystals Using a Harvestable Microfluidic Device. Cryst Growth Des *14*, 3179–3181.

Li, Y.G., and Christie, P.J. (2018). The Agrobacterium VirB/VirD4 T4SS: Mechanism and Architecture Defined Through In Vivo Mutagenesis and Chimeric Systems. Curr. Top. Microbiol. Immunol. *418*, 233–260.

Lin, Y.-H., Lucas, M., Evans, T.R., Abascal-Palacios, G., Doms, A.G., Beauchene, N.A., Rojas, A.L., Hierro, A., and Machner, M.P. (2018). RavN is a member of a previously unrecognized group of Legionella pneumophila E3 ubiquitin ligases. PLoS Pathog. *14*, e1006897.

Lucas, M., Gaspar, A.H., Pallara, C., Rojas, A.L., Fernández-Recio, J., Machner, M.P., and Hierro, A. (2014). Structural basis for the recruitment and activation of the Legionella phospholipase VipD by the host GTPase Rab5. Proc Natl Acad Sci U S A *111*, E3514–E3523.

Luitz, M.P., Bomblies, R., Ramcke, E., Itzen, A., and Zacharias, M. (2016). Adenylylation of Tyr77 stabilizes Rab1b GTPase in an active state: A molecular dynamics simulation analysis. Sci Rep *6*, 19896.

Lüthy, L., Grütter, M.G., and Mittl, P.R.E. (2002). The crystal structure of Helicobacter pylori cysteine-rich protein B reveals a novel fold for a penicillin-binding protein. J. Biol. Chem. *277*, 10187–10193.

Lüthy, L., Grütter, M.G., and Mittl, P.R.E. (2004). The crystal structure of Helicobacter cysteine-rich protein C at 2.0 A resolution: similar peptide-binding sites in TPR and SEL1-like repeat proteins. J. Mol. Biol. *340*, 829–841.

Macchi, P. (2012). Cryo-crystallography: diffraction at low temperature and more. Top Curr Chem *315*, 33–67.

Maguire, C.M., Rösslein, M., Wick, P., and Prina-Mello, A. (2018). Characterisation of particles in solution - a perspective on light scattering and comparative technologies. Sci Technol Adv Mater *19*, 732–745.

Malkin, A.J., Kuznetsov, Yu.G., Glantz, W., and McPherson, A. (1996). Atomic Force Microscopy Studies of Surface Morphology and Growth Kinetics in Thaumatin Crystallization. J. Phys. Chem. *100*, 11736–11743.

Mao, Y., Balkin, D.M., Zoncu, R., Erdmann, K.S., Tomasini, L., Hu, F., Jin, M.M., Hodsdon, M.E., and De Camilli, P. (2009). A PH domain within OCRL bridges clathrin-mediated membrane trafficking to phosphoinositide metabolism. EMBO J. *28*, 1831–1842.

McCoy, A.J. (2017). Acknowledging Errors: Advanced Molecular Replacement with Phaser. Methods Mol. Biol. *1607*, 421–453.

Meir, A., Chetrit, D., Liu, L., Roy, C.R., and Waksman, G. (2018). Legionella DotM structure reveals a role in effector recruiting to the Type 4B secretion system. Nat Commun *9*, 507.

Mellman, I. (1996). Endocytosis and molecular sorting. Annu. Rev. Cell Dev. Biol. *12*, 575–625.

Meng, G., An, X., Ye, S., Liu, Y., Zhu, W., Zhang, R., and Zheng, X. (2013). The crystal structure of LidA, a translocated substrate of the Legionella pneumophila type IV secretion system. Protein Cell *4*, 897–900.

Michard, C., Sperandio, D., Baïlo, N., Pizarro-Cerdá, J., LeClaire, L., Chadeau-Argaud, E., Pombo-Grégoire, I., Hervet, E., Vianney, A., Gilbert, C., et al. (2015). The Legionella Kinase LegK2 Targets the ARP2/3 Complex To Inhibit Actin Nucleation on Phagosomes and Allow Bacterial Evasion of the Late Endocytic Pathway. MBio *6*, e00354-00315.

Mishra, A.K., Del Campo, C.M., Collins, R.E., Roy, C.R., and Lambright, D.G. (2013). The Legionella pneumophila GTPase activating protein LepB accelerates Rab1 deactivation by a non-canonical hydrolytic mechanism. J. Biol. Chem. *288*, 24000–24011.

Mittl, P.R.E., and Schneider-Brachert, W. (2007). Sel1-like repeat proteins in signal transduction. Cell. Signal. *19*, 20–31.

Mizuno-Yamasaki, E., Rivera-Molina, F., and Novick, P. (2012). GTPase networks in membrane traffic. Annu. Rev. Biochem. *81*, 637–659.

Moradi, M., Babin, V., Roland, C., Darden, T.A., and Sagui, C. (2009). Conformations and free energy landscapes of polyproline peptides. Proc. Natl. Acad. Sci. U.S.A. *106*, 20746–20751.

Moreno, A. (2017). Advanced Methods of Protein Crystallization. In Protein Crystallography: Methods and Protocols, A. Wlodawer, Z. Dauter, and M. Jaskolski, eds. (New York, NY: Springer New York), pp. 51–76.

Muder, R.R., and Yu, V.L. (2002). Infection due to Legionella species other than L. pneumophila. Clin. Infect. Dis. *35*, 990–998.

Mukherjee, S., Liu, X., Arasaki, K., McDonough, J., Galán, J.E., and Roy, C.R. (2011). Modulation of Rab GTPase function by a protein phosphocholine transferase. Nature *477*, 103–106.

Müller, M.P., Shkumatov, A.V., Oesterlin, L.K., Schoebel, S., Goody, P.R., Goody, R.S., and Itzen, A. (2012). Characterization of enzymes from Legionella pneumophila involved in reversible adenylylation of Rab1 protein. J. Biol. Chem. *287*, 35036–35046.

Nagai, H., Kagan, J.C., Zhu, X., Kahn, R.A., and Roy, C.R. (2002). A bacterial guanine nucleotide exchange factor activates ARF on Legionella phagosomes. Science *295*, 679–682.

Nair, U., Yen, W.-L., Mari, M., Cao, Y., Xie, Z., Baba, M., Reggiori, F., and Klionsky, D.J. (2012). A role for Atg8-PE deconjugation in autophagosome biogenesis. Autophagy *8*, 780–793.

Nakatogawa, H., Ishii, J., Asai, E., and Ohsumi, Y. (2012). Atg4 recycles inappropriately lipidated Atg8 to promote autophagosome biogenesis. Autophagy *8*, 177–186.

Neunuebel, M.R., Mohammadi, S., Jarnik, M., and Machner, M.P. (2012). Legionella pneumophila LidA affects nucleotide binding and activity of the host GTPase Rab1. J. Bacteriol. *194*, 1389–1400.

Newton, H.J., Sansom, F.M., Bennett-Wood, V., and Hartland, E.L. (2006). Identification of Legionella pneumophila-specific genes by genomic subtractive hybridization with Legionella micdadei and identification of lpnE, a gene required for efficient host cell entry. Infect. Immun. *74*, 1683–1691.

Newton, H.J., Sansom, F.M., Dao, J., McAlister, A.D., Sloan, J., Cianciotto, N.P., and Hartland, E.L. (2007a). Sel1 repeat protein LpnE is a Legionella pneumophila virulence determinant that influences vacuolar trafficking. Infect. Immun. *75*, 5575–5585.

Newton, H.J., Sansom, F.M., Dao, J., McAlister, A.D., Sloan, J., Cianciotto, N.P., and Hartland, E.L. (2007b). Sel1 repeat protein LpnE is a Legionella pneumophila virulence determinant that influences vacuolar trafficking. Infect. Immun. *75*, 5575–5585.

Oliva, G., Sahr, T., and Buchrieser, C. (2018). The Life Cycle of L. pneumophila: Cellular Differentiation Is Linked to Virulence and Metabolism. Front Cell Infect Microbiol *8*, 3.

Olivos-Glander, I.M., Jänne, P.A., and Nussbaum, R.L. (1995). The oculocerebrorenal syndrome gene product is a 105-kD protein localized to the Golgi complex. Am. J. Hum. Genet. *57*, 817–823.

Pankiv, S., Clausen, T.H., Lamark, T., Brech, A., Bruun, J.-A., Outzen, H., Øvervatn, A., Bjørkøy, G., and Johansen, T. (2007). p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. J. Biol. Chem. *282*, 24131–24145.

Patel, C.N., Koh, D.W., Jacobson, M.K., and Oliveira, M.A. (2005). Identification of three critical acidic residues of poly(ADP-ribose) glycohydrolase involved in catalysis: determining the PARG catalytic domain. Biochem. J. *388*, 493–500.

Patterson, A.L. (1934). A Fourier Series Method for the Determination of the Components of Interatomic Distances in Crystals. Phys. Rev. *46*, 372–376.

Perez-Riba, A., Synakewicz, M., and Itzhaki, L.S. (2018). Folding cooperativity and allosteric function in the tandem-repeat protein class. Philos. Trans. R. Soc. Lond., B, Biol. Sci. *373*.

Pike, A.C.W., Garman, E.F., Krojer, T., von Delft, F., and Carpenter, E.P. (2016). An overview of heavy-atom derivatization of protein crystals. Acta Crystallogr D Struct Biol *72*, 303–318.

Plutner, H., Cox, A.D., Pind, S., Khosravi-Far, R., Bourne, J.R., Schwaninger, R., Der, C.J., and Balch, W.E. (1991). Rab1b regulates vesicular transport between the endoplasmic reticulum and successive Golgi compartments. J. Cell Biol. *115*, 31–43.

Poteryaev, D., Datta, S., Ackema, K., Zerial, M., and Spang, A. (2010). Identification of the switch in early-to-late endosome transition. Cell *141*, 497–508.

Powell, H.R. (2017). X-ray data processing. Biosci. Rep. *37*.

Prevost, M.S., and Waksman, G. (2018). X-ray crystal structures of the type IVb secretion system DotB ATPases. Protein Sci. *27*, 1464–1475.

Prevost, M.S., Pinotsis, N., Dumoux, M., Hayward, R.D., and Waksman, G. (2017). The Legionella effector WipB is a translocated Ser/Thr phosphatase that targets the host lysosomal nutrient sensing machinery. Sci Rep *7*, 9450.

Qiu, J., and Luo, Z.-Q. (2017a). Legionella and Coxiella effectors: strength in diversity and activity. Nat. Rev. Microbiol. *15*, 591–605.

Qiu, J., and Luo, Z.-Q. (2017b). Hijacking of the Host Ubiquitin Network by Legionella pneumophila. Front Cell Infect Microbiol *7*, 487.

Quaile, A.T., Stogios, P.J., Egorova, O., Evdokimova, E., Valleau, D., Nocek, B., Kompella, P.S., Peisajovich, S., Yakunin, A.F., Ensminger, A.W., et al. (2018). The Legionella pneumophila effector Ceg4 is a phosphotyrosine phosphatase that attenuates activation of eukaryotic MAPK pathways. J. Biol. Chem. *293*, 3307–3320.

Radaev, S., Li, S., and Sun, P.D. (2006). A survey of protein-protein complex crystallizations. Acta Crystallogr. D Biol. Crystallogr. *62*, 605–612.

Rajan, R., and Matsumura, K. (2018). Development and Application of Cryoprotectants. Adv. Exp. Med. Biol. *1081*, 339–354.

Rajashankar, K., Bufe, A., Weber, W., Eschenburg, S., Lindner, B., and Betzel, C. (2002). Structure of the functional domain of the major grass-pollen allergen Phlp 5b. Acta Crystallogr. D Biol. Crystallogr. *58*, 1175–1181.

Rink, J., Ghigo, E., Kalaidzidis, Y., and Zerial, M. (2005). Rab Conversion as a Mechanism of Progression from Early to Late Endosomes. Cell *122*, 735–749.

Romano-Moreno, M., Rojas, A.L., Williamson, C.D., Gershlick, D.C., Lucas, M., Isupov, M.N., Bonifacino, J.S., Machner, M.P., and Hierro, A. (2017). Molecular mechanism for the subversion of the retromer coat by the Legionella effector RidL. Proc. Natl. Acad. Sci. U.S.A. *114*, E11151–E11160.

Ronau, J.A., and Hochstrasser, M. (2017). The DUB blade goes snicker-snack: Novel ubiquitin cleavage by a Legionella effector protein. Cell Res. *27*, 845–846.

Röntgen, W.C. (1896). ON A NEW KIND OF RAYS. Science *3*, 227–231.

Rupp, B. (2010). Biomolecular Crystallography (Garland Science, Taylor & Francis Group).

Salunke, D.M., and Nair, D.T. (2017). Macromolecular structures: Quality assessment and biological interpretation. IUBMB Life *69*, 563–571.

Saraste, J., Lahtinen, U., and Goud, B. (1995). Localization of the small GTP-binding protein rab1p to early compartments of the secretory pathway. J. Cell. Sci. *108 (Pt 4)*, 1541–1552.

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F.U., and Moarefi, I. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. Cell *101*, 199–210.

Schoebel, S., Cichy, A.L., Goody, R.S., and Itzen, A. (2011). Protein LidA from Legionella is a Rab GTPase supereffector. Proc. Natl. Acad. Sci. U.S.A. *108*, 17945–17950.

Schuller, M., Riedel, K., Gibbs-Seymour, I., Uth, K., Sieg, C., Gehring, A.P., Ahel, I., Bracher, F., Kessler, B.M., Elkins, J.M., et al. (2017). Discovery of a selective allosteric inhibitor targeting macrodomain 2 of poly-adenosine-diphosphate-ribose polymerase 14. ACS Chem Biol *12*, 2866–2874.

Shabalin, I.G., Porebski, P.J., and Minor, W. (2018). Refining the macromolecular model - achieving the best agreement with the data from X-ray diffraction experiment. Crystallogr Rev *24*, 236–262.

Shapiro, A.L., Viñuela, E., and Maizel, J.V. (1967). Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. Biochem. Biophys. Res. Commun. *28*, 815–820.

Sheng, Y., Hong, J.H., Doherty, R., Srikumar, T., Shloush, J., Avvakumov, G.V., Walker, J.R., Xue, S., Neculai, D., Wan, J.W., et al. (2012). A human ubiquitin conjugating enzyme (E2)-HECT E3 ligase structure-function screen. Mol. Cell Proteomics *11*, 329–341.

Shi, X., Halder, P., Yavuz, H., Jahn, R., and Shuman, H.A. (2016). Direct targeting of membrane fusion by SNARE mimicry: Convergent evolution of Legionella effectors. Proc. Natl. Acad. Sci. U.S.A. *113*, 8807–8812.

Shin, H.-W., Hayashi, M., Christoforidis, S., Lacas-Gervais, S., Hoepfner, S., Wenk, M.R., Modregger, J., Uttenweiler-Joseph, S., Wilm, M., Nystuen, A., et al. (2005). An enzymatic cascade of Rab5 effectors regulates phosphoinositide turnover in the endocytic pathway. J. Cell Biol. *170*, 607–618.

Slade, D., Dunstan, M.S., Barkauskaite, E., Weston, R., Lafite, P., Dixon, N., Ahel, M., Leys, D., and Ahel, I. (2011). The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase. Nature *477*, 616–620.

Sledz, P., Zheng, H., Murzyn, K., Chruszcz, M., Zimmerman, M.D., Chordia, M.D., Joachimiak, A., and Minor, W. (2010). New surface contacts formed upon reductive lysine methylation: improving the probability of protein crystallization. Protein Sci. *19*, 1395–1404.

Smith, B.J. (1994). SDS polyacrylamide gel electrophoresis of proteins. Methods Mol. Biol. *32*, 23–34.

Sohn, Y.-S., Shin, H.-C., Park, W.S., Ge, J., Kim, C.-H., Lee, B.L., Heo, W.D., Jung, J.U., Rigden, D.J., and Oh, B.-H. (2015). Lpg0393 of Legionella pneumophila is a guanine-nucleotide exchange factor for Rab5, Rab21 and Rab22. PLoS ONE *10*, e0118683.

Spanò, S., and Galán, J.E. (2018). Taking control: Hijacking of Rab GTPases by intracellular bacterial pathogens. Small GTPases *9*, 182–191.

Steiner, B., Weber, S., and Hilbi, H. (2017). Formation of the Legionella-containing vacuole: phosphoinositide conversion, GTPase modulation and ER dynamics. Int. J. Med. Microbiol.

Stenmark, H., Vitale, G., Ullrich, O., and Zerial, M. (1995). Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion. Cell *83*, 423–432.

Stols, L., Gu, M., Dieckman, L., Raffen, R., Collart, F.R., and Donnelly, M.I. (2002). A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr. Purif. *25*, 8–15.

Stolt, P.C., and Bock, H.H. (2006). Modulation of lipoprotein receptor functions by intracellular adaptor proteins. Cell. Signal. *18*, 1560–1571.

Stoops, E.H., and Caplan, M.J. (2014). Trafficking to the apical and basolateral membranes in polarized epithelial cells. J. Am. Soc. Nephrol. *25*, 1375–1386.

Suh, H.-Y., Lee, D.-W., Lee, K.-H., Ku, B., Choi, S.-J., Woo, J.-S., Kim, Y.-G., and Oh, B.-H. (2010). Structural insights into the dual nucleotide exchange and GDI displacement activity of SidM/DrrA. EMBO J. *29*, 496–504.

Tan, Y., Arnold, R.J., and Luo, Z.-Q. (2011). Legionella pneumophila regulates the small GTPase Rab1 activity by reversible phosphorylcholination. Proc. Natl. Acad. Sci. U.S.A. *108*, 21212–21217.

Thorn, A. (2017). Experimental Phasing: Substructure Solution and Density Modification as Implemented in SHELX. Methods Mol. Biol. *1607*, 357–376.

Toulabi, L., Wu, X., Cheng, Y., and Mao, Y. (2013). Identification and structural characterization of a Legionella phosphoinositide phosphatase. J. Biol. Chem. *288*, 24518–24527.

Trésaugues, L., Silvander, C., Flodin, S., Welin, M., Nyman, T., Gräslund, S., Hammarström, M., Berglund, H., and Nordlund, P. (2014). Structural basis for phosphoinositide substrate recognition, catalysis, and membrane interactions in human inositol polyphosphate 5-phosphatases. Structure *22*, 744–755.

Urbanus, M.L., Quaile, A.T., Stogios, P.J., Morar, M., Rao, C., Di Leo, R., Evdokimova, E., Lam, M., Oatway, C., Cuff, M.E., et al. (2016). Diverse mechanisms of metaeffector activity in an intracellular bacterial pathogen, Legionella pneumophila. Mol. Syst. Biol. *12*, 893.

Urosev, D., Ferrer-Navarro, M., Pastorello, I., Cartocci, E., Costenaro, L., Zhulenkovs, D., Maréchal, J.-D., Leonchiks, A., Reverter, D., Serino, L., et al. (2013). Crystal structure of c5321: a protective antigen present in uropathogenic Escherichia coli strains displaying an SLR fold. BMC Struct. Biol. *13*, 19.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. *66*, 22–25.

Valleau, D., Quaile, A.T., Cui, H., Xu, X., Evdokimova, E., Chang, C., Cuff, M.E., Urbanus, M.L., Houliston, S., Arrowsmith, C.H., et al. (2018). Discovery of Ubiquitin Deamidases in the Pathogenic Arsenal of Legionella pneumophila. Cell Rep *23*, 568–583.

Vieira, O.V., Botelho, R.J., and Grinstein, S. (2002). Phagosome maturation: aging gracefully. Biochem. J. *366*, 689–704.

Walker, J.E. (1998). ATP Synthesis by Rotary Catalysis (Nobel lecture). Angew. Chem. Int. Ed. Engl. *37*, 2308–2319.

Walter, T.S., Meier, C., Assenberg, R., Au, K.-F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. Structure *14*, 1617–1622.

Wang, J., Brudvig, G.W., Batista, V.S., and Moore, P.B. (2017). On the relationship between cumulative correlation coefficients and the quality of crystallographic data sets. Protein Sci. *26*, 2410–2416.

Wang, X., Liang, Y., Fang, Z., Huang, J., and Hong, M. (2019). The intracellular NPxY motif is critical in maintaining the function and expression of human organic anion transporting polypeptide 1B1. Biochim Biophys Acta Biomembr *1861*, 1189–1196.

Watarai, M., Derre, I., Kirby, J., Growney, J.D., Dietrich, W.F., and Isberg, R.R. (2001). Legionella pneumophila is internalized by a macropinocytotic uptake pathway controlled by the Dot/Icm system and the mouse Lgn1 locus. J. Exp. Med. *194*, 1081–1096.

Weber, S.S., Ragaz, C., and Hilbi, H. (2009). The inositol polyphosphate 5-phosphatase OCRL1 restricts intracellular growth of Legionella, localizes to the replicative vacuole and binds to the bacterial effector LpnE. Cell. Microbiol. *11*, 442–460.

Welin, A., Weber, S., and Hilbi, H. (2018). Quantitative Imaging Flow Cytometry of Legionella-Infected Dictyostelium Amoebae Reveals the Impact of Retrograde Trafficking on Pathogen Vacuole Composition. Appl. Environ. Microbiol. *84*.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J. *275*, 1–21.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. FEBS J. *280*, 5705–5736.

Wong, K., Kozlov, G., Zhang, Y., and Gehring, K. (2015). Structure of the Legionella Effector, lpg1496, Suggests a Role in Nucleotide Metabolism. J. Biol. Chem. *290*, 24727–24737.

Wong, K., Perpich, J.D., Kozlov, G., Cygler, M., Abu Kwaik, Y., and Gehring, K. (2017). Structural Mimicry by a Bacterial F Box Effector Hijacks the Host Ubiquitin-Proteasome System. Structure *25*, 376–383.

Xie, Z., Nair, U., and Klionsky, D.J. (2008). Atg8 controls phagophore expansion during autophagosome formation. Mol. Biol. Cell *19*, 3290–3298.

Xu, J., Xu, D., Wan, M., Yin, L., Wang, X., Wu, L., Liu, Y., Liu, X., Zhou, Y., and Zhu, Y. (2017). Structural insights into the roles of the IcmS-IcmW complex in the type IVb secretion system of Legionella pneumophila. Proc. Natl. Acad. Sci. U.S.A. *114*, 13543–13548.

Yang, A., Pantoom, S., and Wu, Y.-W. (2017). Elucidation of the anti-autophagy mechanism of the Legionella effector RavZ using semisynthetic LC3 proteins. Elife *6*.

Yao, J., Yang, F., Sun, X., Wang, S., Gan, N., Liu, Q., Liu, D., Zhang, X., Niu, D., Wei, Y., et al. (2018). Mechanism of inhibition of retromer transport by the bacterial effector RidL. Proc. Natl. Acad. Sci. U.S.A. *115*, E1446–E1454.

Yu, V.L., Plouffe, J.F., Pastoris, M.C., Stout, J.E., Schousboe, M., Widmer, A., Summersgill, J., File, T., Heath, C.M., Paterson, D.L., et al. (2002). Distribution of Legionella Species and Serogroups Isolated by Culture in Patients with Sporadic Community-Acquired Legionellosis: An International Collaborative Survey. J Infect Dis *186*, 127–128.

Zaniew, M., Bökenkamp, A., Kolbuc, M., La Scola, C., Baronio, F., Niemirska, A., Szczepanska, M., Bürger, J., La Manna, A., Miklaszewska, M., et al. (2018). Long-term renal outcome in children with OCRL mutations: retrospective analysis of a large international cohort. Nephrol. Dial. Transplant. *33*, 85–94.

Zhang, N., Gong, X., Lu, M., Chen, X., Qin, X., and Ge, H. (2016). Crystal structures of Apo and GMP bound hypoxanthine-guanine phosphoribosyltransferase from Legionella pneumophila and the implications in gouty arthritis. J. Struct. Biol. *194*, 311–316.

Zhang, X., Jefferson, A.B., Auethavekiat, V., and Majerus, P.W. (1995). The protein deficient in Lowe syndrome is a phosphatidylinositol-4,5-bisphosphate 5-phosphatase. Proc. Natl. Acad. Sci. U.S.A. *92*, 4853–4856.

Zhang, X., Li, X., and Xu, H. (2012). Phosphoinositide isoforms determine compartment-specific ion channel activity. Proc. Natl. Acad. Sci. U.S.A. *109*, 11384–11389.

Zhao, J., Beyrakhova, K., Liu, Y., Alvarez, C.P., Bueler, S.A., Xu, L., Xu, C., Boniecki, M.T., Kanelis, V., Luo, Z.-Q., et al. (2017). Molecular basis for the binding and modulation of V-ATPase by a bacterial effector protein. PLoS Pathog. *13*, e1006394.

Zhu, M., Ma, X., Gao, Y., Li, X., Ke, J., Khan, M.H., Teng, M., Ge, H., Zhu, Z., and Niu, L. (2018). The C-terminus of ubiquitin plays a critical role in deamidase Lpg2148 recognition. Biochem. Biophys. Res. Commun. *503*, 2943–2948.